

CRFE

renewed 5/19/03

Access DB#

93097

5/1 138

5/8 112

# SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name:

David Fox

Examiner #: 65401

Date: 5/4/03

Art Unit: 1638

Phone Number 308-0280

Serial Number: 09/940,550

Mail Box and Bldg/Room Location: CM 19E15 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention:

PLEASE SCAN ONLY

Point of Contact:

Toby Port

Technical Info. Specialist

CM1 6A04

703-308-2521

Inventors (please provide full names):

Earliest Priority Filing Date:

8/00

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

please do SEQUENCE SEARCH

for SEQ ID NOS 9 + 10

(native + mutant form of same)

NO TIMELY HITS

THANK

YOU

- amendment due 5/31 -

## STAFF USE ONLY

### Type of Search

### Vendors and cost where applicable

Searcher:

NA Sequence (#) 2

STN

Searcher Phone #:

AA Sequence (#)

Dialog

Searcher Location:

Structure (#)

Questel/Orbit

Date Searcher Picked Up: 5/6

Bibliographic

Dr.Link

Date Completed: 5/12

Litigation

Lexis/Nexis

Searcher Prep & Review Time: 12

Fulltext

Sequence Systems

Clerical Prep Time:

Patent Family

WWW/Internet

Online Time: 12

Other

Other (specify)

**This Page Blank (uspto)**

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 15:44:18 ; Search time 5977.5 Seconds

(without alignments)  
9888.381 Million cell updates/sec

Title: US-09-940-550A-9

Perfect score: 2031

Sequence: 1 atggcacaagggtgtgtgac.....cggaaagacgtacggtgtag 2031

Scoring table: IDENTITY\_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GeneBlast:

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2031	100.0	2031	6	AX394182	AX394182 Sequence
2	2015	99.2	2031	6	AX394183	AX394183 Sequence
3	2014	99.2	9901	6	AX394184	AX394184 Sequence
4	1812.6	89.2	2200	7	C31INTDNA	X59938 Bacteriophage
5	1812.6	89.2	5711	8	AX114861	AX114861 Sequence
6	1812.6	89.2	5715	12	PCL414670	AJ414670 Cloning v
7	1812.6	89.2	41489	7	BPH6589	AJ006589 Bacterioph
8	200.4	9.9	5967	12	EVU84006	U84006 Expression
9	198.6	9.8	912	6	AX259239	AX259239 Sequence
10	198.6	9.8	1577	6	AX259244	AX259244 Sequence
11	198.6	9.8	1618	6	AX259242	AX259242 Sequence
12	198.6	9.8	2943	6	AX259250	AX259250 Sequence
13	195	9.6	2001	6	AX259240	AX259240 Sequence
14	195	9.6	2001	6	AX356863	AX356863 Sequence
15	195	9.6	2633	6	AR037156	AR037156 Sequence
16	195	9.6	2633	6	AR112042	AR112042 Sequence
17	195	9.6	2725	6	AX259245	AX259245 Sequence
18	195	9.6	2730	6	AX259243	AX259243 Sequence
19	195	9.6	4072	6	AX259251	AX259251 Sequence
20	195	9.6	4341	6	AX259249	AX259249 Sequence
21	195	9.6	5288	12	ASTDNABV	X84105 Artificial
22	195	9.6	5534	6	AR037157	AR037157 Sequence
23	195	9.6	5534	6	AR112043	AR112043 Sequence
24	195	9.6	12817	6	AX356864	AX356864 Sequence
25	195	9.6	13274	6	AX356862	AX356862 Sequence
26	194	9.6	1701	6	AR174855	AR174855 Sequence
27	194	9.6	3336	6	AR174856	AR174856 Sequence
28	194	9.6	3877	6	AR174858	AR174858 Sequence
29	190.2	9.4	5560	6	A60112	A60112 Sequence 5
30	190.2	9.4	5560	6	AR098311	AR098311 Sequence
31	190	9.4	1034	6	A87679	A97679 Sequence 2
32	181.6	8.9	363	6	AX060392	AX060392 Sequence
33	180.6	8.9	363	6	AX060389	AX060389 Sequence
34	180.6	8.9	363	6	AX060391	AX060391 Sequence
35	180.6	8.9	369	6	AX060390	AX060390 Sequence
36	179.8	8.9	3699	8	STLSIG	X04753 Potato ligh
37	64	3.2	125020	9	AF429315	AF429315 Homo sapi
38	63.8	3.1	39314	1	SGR300302	AJ300302 Streptomy
39	60.4	3.0	125020	9	AF429315	AF429315 Homo sapi
40	59	2.9	134375	2	AC118169	AC118169 Rattus no
41	58.6	2.9	169546	2	AC004157	AC004157 Plasmodiu
42	58.2	2.9	154902	9	AC005799	AC005799 Homo sapi
43	58.2	2.9	191947	2	AC079210	AC079210 Homo sapi
44	58	2.9	9424	8	SDO404228	AJ404228 Saccharom
45	57.8	2.8	1181	8	MISCOX13	X14910 Yeast mitoc

ALIGNMENTS

RESULT 1  
AX394182  
LOCUS AX394182 2031 bp DNA linear PAT 23-MAR-2002  
DEFINITION Sequence 9 from Patent WO0216609.  
ACCESSION AX394182  
VERSION AX394182.1 GI:19702116  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS Mankin, L.  
TITLE Self-excising polynucleotides and uses thereof  
JOURNAL Patent: WO 0216609-A 9 28-FEB-2002.  
BASF Plant Science GmbH (DE) ; McKersie, Bryan (US)

filed at US  
instat US

FEATURES  
source  
Location/Qualifiers  
1. 2031  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="C31int DNA sequence"  
BASE COUNT 451 a 562 c 640 g 378 t  
ORIGIN

Query Match 100.0%; Score 2031; DB 6; Length 2031;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2031; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

-QY 1 ATGGCACAAGGGGTTGTGACCGGGGTGATACGTAACTTCTGCTTCTACCTTTGATATA 60  
Db 1 ATGGCACAAGGGGTTGTGACCGGGGTGATACGTAACTTCTGCTTCTACCTTTGATATA 60  
QY 61 TAT 120  
Db 61 TAT 120  
QY 121 AAAGATGTAGTATATAGCAATGCTTTCTAGTTTATATAGTGTATATTTTATTT 180  
Db 121 AAAGATGTAGTATATAGCAATGCTTTCTAGTTTATATAGTGTATATTTTATTT 180  
QY 181 ATACTTTTCTAT 240  
Db 181 ATACTTTTCTAT 240  
QY 241 CGTCACTGCGCGGAGCGGAGAAATTCGACGCGCAGCAAGCCGACACAGCGTACGAC 300  
Db 241 CGTCACTGCGCGGAGCGGAGAAATTCGACGCGCAGCAAGCCGACACAGCGTACGAC 300  
QY 301 AACGAAGACAAGCGCGCGACCTTCAGCGCGGAGTTCAGCGCGGAGCGCGGCGTT 360  
Db 301 AACGAAGACAAGCGCGCGACCTTCAGCGCGGAGTTCAGCGCGGAGCGCGGCGTT 360  
QY 361 TTGCTGCGGCAATTCAGCGAAGCGCGCGCAGCTCGGCGTTGCGGAGCGCGCGCG 420  
Db 361 TTGCTGCGGCAATTCAGCGAAGCGCGCGCAGCTCGGCGTTGCGGAGCGCGCGCG 420  
QY 421 GAGTTCGAAGCGCATCTGAACGAATGCGCGCGCGCGCGCTCAACATGATCTAT 480  
Db 421 GAGTTCGAAGCGCATCTGAACGAATGCGCGCGCGCGCGCTCAACATGATCTAT 480  
QY 481 GAGTTCGCGGCTTCTCGCGCTGAAGTTCATGACGCGCATTCGATTTCTCGGAAT 540  
Db 481 GAGTTCGCGGCTTCTCGCGCTGAAGTTCATGACGCGCATTCGATTTCTCGGAAT 540  
QY 541 CTGCGCTGCGGCTGACGATTTCTCCACTCAGGAAGCGCTCTCCGCGCAGGAAAC 600  
Db 541 CTGCGCTGCGGCTGACGATTTCTCCACTCAGGAAGCGCTCTCCGCGCAGGAAAC 600  
QY 601 ATGACCTGATTCACCTGATTTATGCGGCTGACGCGCTCGCACAAAGATCTTCG 660  
Db 601 ATGACCTGATTCACCTGATTTATGCGGCTGACGCGCTCGCACAAAGATCTTCG 660  
QY 661 TCGGCGAAGATTTTCGACACGAAAGACCTTCAGCGGCAATTCGCGGCGGTA 720  
Db 661 TCGGCGAAGATTTTCGACACGAAAGACCTTCAGCGGCAATTCGCGGCGGTA 720  
QY 721 AAGGCGCTTACGCGCTTCGAGCTTTTCGAGAGAGAGAGATCAGCGCGCAACG 780  
Db 721 AAGGCGCTTACGCGCTTCGAGCTTTTCGAGAGAGAGAGATCAGCGCGCAACG 780  
QY 781 ATGTCATATGCTCATCAACAAGCTTCGCGCACTCCCTTACCGGACCTTC 840  
Db 781 ATGTCATATGCTCATCAACAAGCTTCGCGCACTCCCTTACCGGACCTTC 840  
QY 841 GAGTTCAGCGCGGAGCTATCCGCTGCTGCTGAGATCAAGACGCAAAACACTT 900  
Db 841 GAGTTCAGCGCGGAGCTATCCGCTGCTGCTGAGATCAAGACGCAAAACACTT 900  
QY 901 CCTTCAAGCGCGGAGCTACAGCGCGGAGCATCAGCGGCGCTTTGTAAG 960

Db 901 CCTTCAAGCGCGGAGCTACAGCGCGGAGCATCAGCGGCGCTTTGTAAG 960  
QY 961 CGCATGACGCTGACGCGCGTGCAGCGCGCGGCGGAGAGATTTGGAAGAACCGCTTCA 1020  
Db 961 CGCATGACGCTGACGCGCGTGCAGCGCGCGGCGGAGAGATTTGGAAGAACCGCTTCA 1020  
QY 1021 AGCGCTGCGGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1080  
Db 1021 AGCGCTGCGGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1080  
QY 1081 GCGGCTGAGGTGATCTACAGAAGAAGCGCGGAGCGGAGCGGAGCGGAGCGGAG 1140  
Db 1081 GCGGCTGAGGTGATCTACAGAAGAAGCGCGGAGCGGAGCGGAGCGGAGCGGAG 1140  
QY 1141 TACCGCATTCAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1200  
Db 1141 TACCGCATTCAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1200  
QY 1201 ATCGAGCGCGCTGAGGTGATGAGCTTCAGCGGCGGAGCGGAGCGGAGCGGAG 1260  
Db 1201 ATCGAGCGCGCTGAGGTGATGAGCTTCAGCGGCGGAGCGGAGCGGAGCGGAG 1260  
QY 1261 GCGCTTCCCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1320  
Db 1261 GCGCTTCCCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1320  
QY 1321 GCGCTCATGCTTCGAAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 1380  
Db 1321 GCGCTCATGCTTCGAAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 1380  
QY 1381 CGGAAGGTGCTGACCGCGCTCCGCACTGGGAGCAGCAGGAGCGGAGCGGAGCG 1440  
Db 1381 CGGAAGGTGCTGACCGCGCTCCGCACTGGGAGCAGCAGGAGCGGAGCGGAGCG 1440  
QY 1441 GCGGACCTGCAACAAGTCTGCTGCGGAACGCTTCACCAAGATCAGGAGCGGAG 1500  
Db 1441 GCGGACCTGCAACAAGTCTGCTGCGGAACGCTTCACCAAGATCAGGAGCGGAG 1500  
QY 1501 GACGAAGAGAGCTTGGCGCTTCTGTTGGAAGCGCGCGCGCGCGCGCGCG 1560  
Db 1501 GACGAAGAGAGCTTGGCGCTTCTGTTGGAAGCGCGCGCGCGCGCGCGCGCG 1560  
QY 1561 GCGCTGAGAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1620  
Db 1561 GCGCTGAGAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1620  
QY 1621 GCGCTTGAAGAGCTGTAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680  
Db 1621 GCGCTTGAAGAGCTGTAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680  
QY 1681 AAGCACTTCCGGAAGCAAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1740  
Db 1681 AAGCACTTCCGGAAGCAAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1740  
QY 1741 CTGCGGACTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1800  
Db 1741 CTGCGGACTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1800  
QY 1801 GAGCGGAGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1860  
Db 1801 GAGCGGAGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1860  
QY 1861 AAGCGCTGCTGCTGCGGCTCTTCTGAGCAAGATCGTTGTCACGAGTTCAG 1920  
Db 1861 AAGCGCTGCTGCTGCGGCTCTTCTGAGCAAGATCGTTGTCACGAGTTCAG 1920  
QY 1921 AGGGGCGAGGGAAGCGCGCATCGAGAAGCGCGCTTCGATCAGTGGGAGCGCG 1980  
Db 1921 AGGGGCGAGGGAAGCGCGCATCGAGAAGCGCGCTTCGATCAGTGGGAGCGCG 1980  
QY 1981 GACGAGCAGGAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2031





QY 1741 CTGCGCAACTTGAAGCCCGGAGCCCGCAAGCTTCCCTTGACCAATGTTCCCGAA 1800  
Db 1741 CTTGCCGAACCTTGAGCCCGCGGAGCCCGCAAGTTGCCCTTGACCAATGTTCCCGAA 1800  
QY 1801 GACGCCGACGCTGACCCGACCGGCCCTTAAGTCGTGGTGGGGCGCGCTCAGTAGACGAC 1860  
Db 1801 GACGCCGACGCTGACCCGACCGGCCCTTAAGTCGTGGTGGGGCGCGCTCAGTAGACGAC 1860  
QY 1861 AAGCGCGTGTTCGTGGGCTCTTCTGTAGACAAGATCGTGTACGAGTCGACTACGGGC 1920  
Db 1861 AAGCGCGTGTTCGTGGGCTCTTCTGTAGACAAGATCGTGTACGAGTCGACTACGGGC 1920  
-QY 1921 AGGGGGCAGGGAACGCCCATCGAGAACCGCGCTTCGATCAGTGGCGAAGCCGCCGACC 1980  
Db 1921 AGGGGGCAGGGAACGCCCATCGAGAACCGCGCTTCGATCAGTGGCGAAGCCGCCGACC 1980  
QY 1981 GACGACGACGAGACGACGCCCGGACGCGGACGGAAGCGTAGCGGCTAG 2031  
Db 1981 GACGACGACGAGACGACGCCCGGACGCGGACGGAAGCGTAGCGGCTAG 2031

RESULT 3  
LOCUS AX394184 9901 bp DNA linear PAT 23-MAR-2002  
DEFINITION Sequence 11 from Patent WO0216609.  
ACCESSION AX394184  
VERSION AX394184.1 GI:19702118  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
REFERENCE 1  
AUTHORS Mankin, L.  
TITLE Self-excising polynucleotides and uses thereof  
JOURNAL Patent: WO 0216609-A 11 28-FEB-2002;  
FEATURES  
source location/Qualifiers  
1. .9901  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="PBPS EM051 T-DNA region"  
BASE COUNT 2489 a 2371 c 2478 g 2551 t 12 others  
ORIGIN

Query Match 99.2%; Score 2014; DB 6; Length 9901;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2031; Conservative 0; Mismatches 0; Indels 7; Gaps 1;

QY 1 ATGGCACAAGGGGTGTGACCGGGGTGATAGTATGCTTCTACCTTGATATA 60  
Db 4984 ATGGCACAAGGGGTGTGACCGGGGTGATAGTATGCTTCTACCTTGATATA 5043  
QY 61 TATATAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120  
Db 5044 TATATAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5103  
QY 121 AAAGAATGTAGTATATAGCAATGCTTCTGTAGTTAATAGTGTATATTTAATT 180  
Db 5104 AAAGAATGTAGTATATAGCAATGCTTCTGTAGTTAATAGTGTATATTTAATT 5163  
QY 181 ATAACCTTTCTAATATATGACCAAAATTTGTGATGTGACGAGTACGGCGGTGCTTACGAC 240  
Db 5164 ATAACCTTTCTAATATATGACCAAAATTTGTGATGTGACGAGTACGGCGGTGCTTACGAC 5223  
QY 241 CGTC-----AGTCGGCGGAGCGGAGAAATTCGAGCGCAGCAAGCCGACACAGCG 293  
Db 5224 CGTCGRNTNDAGTCGGCGGAGCGGAGAAATTCGAGCGCAGCAAGCCGACACAGCG 5283  
QY 294 TAGCGCCACAAGACAAAGCGCGGACCTTCAGCGCGAAGTGAAGCGGAGCGGCGG 353  
Db 5284 TAGCGCCACAAGACAAAGCGCGGACCTTCAGCGCGAAGTGAAGCGGAGCGGCGG 5343  
QY 354 GTTCAGGTTGTCGGGCAATTCAGCGAAGCGCGGCAAGTGTGGCGGAGCGGCGGA 413

Db 5344 GTTCAGGTTGTCGGGCAATTCAGCGAAGCGCGGCAAGTGTGGCGGAGCGGCGGA 5403  
QY 414 GCGCCGAGGTTGCAAGGCATCTGAAAGAAATGCGCGCGCGCGCTCAACATGATCAT 473  
Db 5404 GCGCCGAGGTTGCAAGGCATCTGAAAGAAATGCGCGCGCGCGCTCAACATGATCAT 5463  
QY 474 TGTCTATGACGTGTGCGGCTTCTCGCGCTGAAGGTATGACGACCGCATTCGATTTGCTC 533  
Db 5464 TGTCTATGACGTGTGCGGCTTCTCGCGCTGAAGGTATGACGACCGCATTCGATTTGCTC 5523  
QY 534 GGAATTGCTCGCCCTGGGCGGTGACGATTTGTTCCACTCAGGAAGCGCTTCCGCGAGGG 593  
Db 5524 GGAATTGCTCGCCCTGGGCGGTGACGATTTGTTCCACTCAGGAAGCGCTTCCGCGAGGG 5583  
QY 594 AAAGTCATGACCTGATTCACCTGATTAATGCGGCTCGACGCGTCCACAAAGAAATCTTC 653  
Db 5584 AAAGTCATGACCTGATTCACCTGATTAATGCGGCTCGACGCGTCCACAAAGAAATCTTC 5643  
QY 654 GCTGAAGTCGCGCAAGATTCTCGACACGAAAGACCTTCAGCGCGAATTGGCGGCTACGT 713  
Db 5644 GCTGAAGTCGCGCAAGATTCTCGACACGAAAGACCTTCAGCGCGAATTGGCGGCTACGT 5703  
QY 714 CGCGGGGAAGCGCCTTACGCGCTTCGAGCTTGTTCGAGACGAAAGAGATCAGCGCAA 773  
Db 5704 CGCGGGGAAGCGCCTTACGCGCTTCGAGCTTGTTCGAGACGAAAGAGATCAGCGCAA 5763  
QY 774 CGGCGGAATGTCATATGTCGTATCAACAAGCTTGGCGACTCGACCACTCCCTTACCGG 833  
Db 5764 CGGCGGAATGTCATATGTCGTATCAACAAGCTTGGCGACTCGACCACTCCCTTACCGG 5823  
QY 834 ACCCTTCGAGTTCGAGCCCGGACGTAATCCGGTGTGGTGGCTGAGATCAAGACGCACAA 893  
Db 5824 ACCCTTCGAGTTCGAGCCCGGACGTAATCCGGTGTGGTGGCTGAGATCAAGACGCACAA 5883  
QY 894 ACACCTTCCTTCAAGCCGCGGAGTCAAGCCGCTTCAACCCGGGCGACGATCACGGGGCT 953  
Db 5884 ACACCTTCCTTCAAGCCGCGGAGTCAAGCCGCTTCAACCCGGGCGACGATCACGGGGCT 5943  
QY 954 TTGTAAGCGCATGACGCTGACGCGCGTGGCGACCGGGGCGAGACGATGGGAAGAAGAC 1013  
Db 5944 TTGTAAGCGCATGACGCTGACGCGCGTGGCGACCGGGGCGAGACGATGGGAAGAAGAC 6003  
QY 1014 CGCTTCAGCGCCTGGAGCCCGGCAACCGTTATGCGAATCTTCGGGACCCGCTATTGC 1073  
Db 6004 CGCTTCAGCGCCTGGAGCCCGGCAACCGTTATGCGAATCTTCGGGACCCGCTATTGC 6063  
QY 1074 GGGCTTCGCGGCTGAGGTGATCTCAAGAAGAAGCCGGACGGCACCGGCAAGAGAT 1133  
Db 6064 GGGCTTCGCGGCTGAGGTGATCTCAAGAAGAAGCCGGACGGCACCGGCAAGAGAT 6123  
QY 1134 TGAGGGTTACCGCATTCAGCGCGGACCCGATCAGCTCCGGCGGTGAGCTTGATTGCGG 1193  
Db 6124 TGAGGGTTACCGCATTCAGCGCGGACCCGATCAGCTCCGGCGGTGAGCTTGATTGCGG 6183  
QY 1194 ACCGATCATGAGCCCGCTGAGTGTATGAGCTTCAGGCGTGTGGAGCGGAGGGGCG 1253  
Db 6184 ACCGATCATGAGCCCGCTGAGTGTATGAGCTTCAGGCGTGTGGAGCGGAGGGGCG 6243  
QY 1254 CGGCAAGGGGCTTCCCGGGGCAAGCCATCTGTCCGCGCATGACAAAGCTGTACTGCGA 1313  
Db 6244 CGGCAAGGGGCTTCCCGGGGCAAGCCATCTGTCCGCGCATGACAAAGCTGTACTGCGA 6303  
QY 1314 GTGTGGCGCGCTCAAGACTTCGAAGCGGCGGGAAGAATCATCAAGGACTCTTACCGCTG 1373  
Db 6304 GTGTGGCGCGCTCAAGACTTCGAAGCGGCGGGAAGAATCATCAAGGACTCTTACCGCTG 6363  
QY 1374 CCGTCGCGGGAAGGTGTCGACCCGTCGCACTGGGCGAGCAGAAAGCAGTGCAGAGT 1433  
Db 6364 CCGTCGCGGGAAGGTGTCGACCCGTCGCACTGGGCGAGCAGAAAGCAGTGCAGAGT 6423  
QY 1434 CAGCATGGCGGCACTGACAAAGTTGTCGGGAAGCGCATCTTCAACAAGATCAGCGACGC 1493



QY 693 GCCGAATTGGCGGGTACGTGCGGGGGAAGGCCCTTACGGCTTCGAGCTTGTTCGGA 752  
 Db 735 GCGGAATTGGCGGGTACGTGCGGGGGAAGGCCCTTACGGCTTCGAGCTTGTTCGGA 794  
 QY 753 GACGAAGAGATCACGGCGCAACGGCCGAATGTCAATGTCTCATCAACAAGCTTGCGCA 812  
 Db 795 GACGAAGAGATCACGGCGCAACGGCCGAATGTCAATGTCTCATCAACAAGCTTGCGCA 854  
 QY 813 CTCGACCACTCCCTTACCGGACCCCTTCGAGCTTCGAGCCCGACGTAATCCGGTGTGTG 872  
 Db 855 CTCGACCACTCCCTTACCGGACCCCTTCGAGCTTCGAGCCCGACGTAATCCGGTGTGTG 914  
 QY 873 GCGTGAGATCAAGACCCACAACACCTTCCCTTCAAGCCGGGAGTCAAGCCGCATTCGA 932  
 Db 915 GCGTGAGATCAAGACCCACAACACCTTCCCTTCAAGCCGGGAGTCAAGCCGCATTCGA 974  
 QY 933 CCGGGGACGATCACGGGGCTTGTAAAGCGCATGAGCGCTGACGCCGTGCCAGCCGGGG 992  
 Db 975 CCGGGGACGATCACGGGGCTTGTAAAGCGCATGAGCGCTGACGCCGTGCCAGCCGGGG 1034  
 QY 993 CGAGACGATTGGGAAGAAGACCGCTTCAAGCCGCTGGAGCCCGGCAACCGTTATGGCAAT 1052  
 Db 1035 CGAGACGATTGGGAAGAAGACCGCTTCAAGCCGCTGGAGCCCGGCAACCGTTATGGCAAT 1094  
 QY 1053 CCTTCGGACCCGCTATTGCGGGCTTCGCCGCTGAGGTGATCTACAAGAAGACCCGGA 1112  
 Db 1095 CCTTCGGACCCGCTATTGCGGGCTTCGCCGCTGAGGTGATCTACAAGAAGACCCGGA 1154  
 QY 1113 CGGACGCGCCGACCAAGAGATTGAGGGTTACCGCATTCAGCCGCGACCCGATCACGCTCG 1172  
 Db 1155 CGGACGCGCCGACCAAGAGATTGAGGGTTACCGCATTCAGCCGCGACCCGATCACGCTCG 1214  
 QY 1173 GCCGCTGAGCTTGATTGCGGACCATCATCGAGCCCGCTGAGTGTATGAGCTTCAGGC 1232  
 Db 1215 GCCGCTGAGCTTGATTGCGGACCATCATCGAGCCCGCTGAGTGTATGAGCTTCAGGC 1274  
 QY 1233 GTGTTGAGCGGACGAGGGGGCGGCAAGGGGCTTCCCGGGGCAAGCCATTCGTCCGC 1292  
 Db 1275 GTGTTGAGCGGACGAGGGGGCGGCAAGGGGCTTCCCGGGGCAAGCCATTCGTCCGC 1334  
 QY 1293 CATGACAACTGTACTGCGAGTGTGGCGCCGCTCATGACTTCGAAGCGCGGGAAGAAATC 1352  
 Db 1335 CATGACAACTGTACTGCGAGTGTGGCGCCGCTCATGACTTCGAAGCGCGGGAAGAAATC 1394  
 QY 1353 GATCAAGACTCTTACCGCTGCGCTGCGCGGAAGGTGTGACCCGCTCCGCACTGGGCA 1412  
 Db 1395 GATCAAGACTCTTACCGCTGCGCTGCGCGGAAGGTGTGACCCGCTCCGCACTGGGCA 1454  
 QY 1413 GCACGAAGGACGTCAGACGTCAGCATGGCGGCACTCGACAAGTTCGTTGCCGAACGCAT 1472  
 Db 1455 GCACGAAGGACGTCAGACGTCAGCATGGCGGCACTCGACAAGTTCGTTGCCGAACGCAT 1514  
 QY 1473 CTTCAACAAGATCAGGACGCGCGGAAGGCGAGCAAGACGTTGGCGCTTCTGTGGAGC 1532  
 Db 1515 CTTCAACAAGATCAGGACGCGCGGAAGGCGAGCAAGACGTTGGCGCTTCTGTGGAGC 1574  
 QY 1533 CGCCCGACGCTTCGCAAGCTCTAGAGGCGCTGAGAAGAGCGGGAACGGGCAACCT 1592  
 Db 1575 CGCCCGACGCTTCGCAAGCTCTAGAGGCGCTGAGAAGAGCGGGAACGGGCAACCT 1634  
 QY 1593 TGTTCGGAGCGCGCGGACGCGCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGC 1652  
 Db 1635 TGTTCGGAGCGCGCGGACGCGCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGC 1694  
 QY 1653 AGCGCGCTACGACGACCCGCTTGGCAGAGCACTTCGGAAGCAACAGGCGAGCTGAC 1712  
 Db 1695 AGCGCGCTACGACGACCCGCTTGGCAGAGCACTTCGGAAGCAACAGGCGAGCTGAC 1754  
 QY 1713 GCTCCGCGACGAAGGCGGGAAGAGCGGCTTCGCAACTTGAAGCCGCGGAAGCCCGAA 1772  
 Db 1755 GCTCCGCGACGAAGGCGGGAAGAGCGGCTTCGCAACTTGAAGCCGCGGAAGCCCGAA 1814  
 QY 1773 GCTTCCCTTGAACAAATGTTCCCGAAGACGCGAGCTGACCCGACCGGCTTAAGTC 1832

Db 1815 GCTTCCCTTGACCAATGTTCCCGAAGACGCGCGAGCTGACCCGAGCCCTTAAGTC 1874  
 QY 1833 GTGTGGGGCGCGCGTACAGTACGACGACGAAGCGCGGTTCGTCGGGCTTCGTAGACAA 1892  
 Db 1875 GTGTGGGGCGCGCGTACAGTACGACGACGAAGCGCGGTTCGTCGGGCTTCGTAGACAA 1934  
 QY 1893 GATGTTGTACGAAGTGCAGTACGAGGGGAGGGGAGGGAACGCCCATCGAAGACCGCGC 1952  
 Db 1935 GATGTTGTACGAAGTGCAGTACGAGGGGAGGGGAGGGAACGCCCATCGAAGACCGCGC 1994  
 QY 1953 TTGATCAGTGGGGGGAAGCGCGCGACCGACGACGACGAAGACGACGCGCCAGAGCGGAC 2012  
 Db 1995 TTGATCAGTGGGGGGAAGCGCGCGACCGACGACGACGAAGACGACGCGCCAGAGCGGAC 2054  
 QY 2013 GGAAGACGTAGCGGCTAG 2031  
 Db 2055 GGAAGACGTAGCGGCTAG 2073

RESULT 5  
 AX114861  
 LOCUS AX114861 5711 bp DNA linear PAT 11-MAY-2001  
 DEFINITION Sequence 21 from Patent WO0129208.  
 ACCESSION AX114861  
 VERSION AX114861.1 GI:14031803  
 KEYWORDS  
 SOURCE synthetic construct.  
 ORGANISM synthetic construct.  
 REFERENCE 1 (bases 1 to 5711)  
 AUTHORS Kuehn, R., von Melchner, H. and Altschmied, J.  
 TITLE Conditional gene trapping construct for the disruption of genes  
 JOURNAL Patent: WO 0129208-A 21 26-APR-2001;  
 ARTEMIS pharmaceuticals GmbH (DE) ; Frankgen Biotechnologie AG (DE)  
 FEATURES  
 Source  
 1. 5711  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="vector pRK65"  
 BASE COUNT 1329 a 1511 c 1617 g 1254 t  
 ORIGIN  
 Query Match 89.2%; Score 1812.6; DB 6; Length 5711;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1815; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 213 GATGTGACGTACGCGGGGTGCTTACGACCGTCACTGCGCGGAGCGCGGAGAAATTCAGCGC 272  
 Db 1001 GGTGACACGTACGCGGGGTGCTTACGACCGTCACTGCGCGGAGCGCGGAGAAATTCAGCGC 1060  
 QY 273 AGCAAGCCACGACACAGCGGTACGCGCCCAAGGAAGAGAGGGCGCGACCTTCAGCGCGA 332  
 Db 1061 AGCAAGCCACGACACAGCGGTACGCGCCCAAGGAAGAGAGGGCGCGACCTTCAGCGCGA 1120  
 QY 333 AGTCAGCGCGACGCGGGCGGCTTACGCTGCTCGGGCATTCAGGGAAGCGCGCGGAC 392  
 Db 1121 AGTCAGCGCGACGCGGGCGGCTTACGCTGCTCGGGCATTCAGGGAAGCGCGGCGAC 1180  
 QY 393 GTCGGCTTCGGGACGCGCGGAGCGCGGAGTTGCAACGCATCCTGAACGAATGCCGCGC 452  
 Db 1181 GTCGGCTTCGGGACGCGCGGAGCGCGGAGTTGCAACGCATCCTGAACGAATGCCGCGC 1240  
 QY 453 CGGGCGGCTCAACATGATGATGCTATGACGCTGCGGCTTCTCGGCTGAAGTCAT 512  
 Db 1241 CGGGCGGCTCAACATGATGATGCTATGACGCTGCGGCTTCTCGGCTGAAGTCAT 1300  
 QY 513 GGACGGAATTCGATGCTTCGGAATTGCTCGCCCTGGGCGTGACGATGTTTCCACTCA 572  
 Db 1301 GGACGGAATTCGATGCTTCGGAATTGCTCGCCCTGGGCGTGACGATGTTTCCACTCA 1360  
 QY 573 GGAAGCGCTTCCGCGAAGGAACGTATGACCTGATTCACCTGATTAATGCGGCTGA 632



Db 1361 GGAAGGCGTCTTCCGGCAGGGAACGTCATGACCTGATTCACCTGATTTATCGGGCTCGA 1420  
QY 633 CCGGTGACACAAGAATCTTCGCTGAAGTCGGCGAAGATTCTCGACAGAGAACCTTCA 692  
Db 1421 CCGGTGACACAAGAATCTTCGCTGAAGTCGGCGAAGATTCTCGACAGAGAACCTTCA 1480  
QY 693 GCGCAATTGGCGGGTACGTCGGCGGGAAGGCGCTTACGGCTTCGAGCTTGTTCGGA 752  
Db 1481 GCGCAATTGGCGGGTACGTCGGCGGGAAGGCGCTTACGGCTTCGAGCTTGTTCGGA 1540  
QY 753 GACGAAGAGATCACGCGCAACGGCCGAATGTCATATGTCATCAACAGCTTGGCA 812  
Db 1541 GACGAAGAGATCACGCGCAACGGCCGAATGTCATATGTCATCAACAGCTTGGCA 1600  
QY 813 CTCGACCACTCCCTTACGGGACCCCTTCGAGTTCGAGCCCGACGTAATCCGCTGGTG 872  
Db 1601 CTCGACCACTCCCTTACGGGACCCCTTCGAGTTCGAGCCCGACGTAATCCGCTGGTG 1660  
QY 873 GCGTAGATCAAGACGACAAACACCTTCCCTTCAAGCCGGGAGTCAAGCCGCTTCA 932  
Db 1661 GCGTAGATCAAGACGACAAACACCTTCCCTTCAAGCCGGGAGTCAAGCCGCTTCA 1720  
QY 933 CCGCGGACGATCACGGGCTTGTAGCGCATGACGCTGACGCGCTGCGGACCCGGG 992  
Db 1721 CCGCGGACGATCACGGGCTTGTAGCGCATGACGCTGACGCGCTGCGGACCCGGG 1780  
QY 993 CGAGACGATTGGGAAGAACCGCTTCAAGCGGCTGGGACCCGCAACCGTTATGCGAAT 1052  
Db 1781 CGAGACGATTGGGAAGAACCGCTTCAAGCGGCTGGGACCCGCAACCGTTATGCGAAT 1840  
QY 1053 CATTTCGGGACCCCGCTATTCGGGGCTTCGCGCTGAGTGATCTACAAGAGAAGCCGGA 1112  
Db 1841 CATTTCGGGACCCCGCTATTCGGGGCTTCGCGCTGAGTGATCTACAAGAGAAGCCGGA 1900  
QY 1113 CCGCACGCGGACCGACGAGATTGAGGGTTACCGCATTCAGCGGACCCGATCAGCTCCG 1172  
Db 1901 CCGCACGCGGACCGACGAGATTGAGGGTTACCGCATTCAGCGGACCCGATCAGCTCCG 1960  
QY 1173 GCCGGTCGAGCTTGATTGCGGACCGATCATCGAGCCGCTGAGTGATGAGCTTCAGGC 1232  
Db 1961 GCCGGTCGAGCTTGATTGCGGACCGATCATCGAGCCGCTGAGTGATGAGCTTCAGGC 2020  
QY 1233 GTGCTTGACGCGGAGGGGCGGCAAGGGGCTTCCCGGGGCAAGCCATCTGTCCGC 1292  
Db 2021 GTGCTTGACGCGGAGGGGCGGCAAGGGGCTTCCCGGGGCAAGCCATCTGTCCGC 2080  
QY 1293 CATGACACAAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAGCGCGGGAAGATC 1352  
Db 2081 CATGACACAAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAGCGCGGGAAGATC 2140  
QY 1353 GATCAAGGACTCTTACCGCTGCCGTGCCGGAAGGTGTCGACCCGTCGCACTGGGCA 1412  
Db 2141 GATCAAGGACTCTTACCGCTGCCGTGCCGGAAGGTGTCGACCCGTCGCACTGGGCA 2200  
QY 1413 GCACGAAGGACGTCACGTCAGCATGGCGGCACTCGACAAGTTGCTGGGAAGCAT 1472  
Db 2201 GCACGAAGGACGTCACGTCAGCATGGCGGCACTCGACAAGTTGCTGGGAAGCAT 2260  
QY 1473 CTTCAACAAGATCAGGACGCGCGAAGGCGAGCAAGACGTTGGCGCTTCTGTGGGAAGC 1532  
Db 2261 CTTCAACAAGATCAGGACGCGCGAAGGCGAGCAAGACGTTGGCGCTTCTGTGGGAAGC 2320  
QY 1533 CGCCCGACGCTTCGGAAGCTCACTGAGCGCGCTGAGAAGAGCGGGAACCGCAACCT 1592  
Db 2321 CGCCCGACGCTTCGGAAGCTCACTGAGCGCGCTGAGAAGAGCGGGAACCGCAACCT 2380  
QY 1593 TGTTCGCGAGCGCGCGCTGAACGCGCTTGAAGAGCTGTACGAAGACCGCGCGGC 1652  
Db 2381 TGTTCGCGAGCGCGCGCTGAACGCGCTTGAAGAGCTGTACGAAGACCGCGCGGC 2440  
QY 1653 AGCGCGCTAGCAGCGGACCGCTTGGCAGGAAGCACTTCCGGAAGCAACAGCGCGCTGAC 1712  
Db 2441 AGCGCGCTAGCAGCGGACCGCTTGGCAGGAAGCACTTCCGGAAGCAACAGCGCGCTGAC 2500

QY 1713 GCTCCGGCAGCAAGGGGCGGAAGAGCGGCTTGCCGAACTTGAAGCCGCGAAGCCCGAA 1772  
Db 2501 GCTCCGGCAGCAAGGGGCGGAAGAGCGGCTTGCCGAACTTGAAGCCGCGAAGCCCGAA 2560  
QY 1773 GCTTCCCTTGACCAATGTTTCCCGAAGACGCGGACGCTGACCCGACCGGCTTAAGTC 1832  
Db 2561 GCTTCCCTTGACCAATGTTTCCCGAAGACGCGGACGCTGACCCGACCGGCTTAAGTC 2620  
QY 1833 GTGTGGGCGCGCGCTCAGTAGACGACAAGCGGCTGTTCGTGGGCTCTTCTGAGACAA 1892  
Db 2621 GTGTGGGCGCGCGCTCAGTAGACGACAAGCGGCTGTTCGTGGGCTCTTCTGAGACAA 2680  
QY 1893 GATCGTTGTCACGAAGTGCATACGGGCGAGGGGCGAGGAAGCCCATCGAGAAGCGCGC 1952  
Db 2681 GATCGTTGTCACGAAGTGCATACGGGCGAGGGGCGAGGAAGCCCATCGAGAAGCGCGC 2740  
QY 1953 TTCGATCAGCTGGGCGAAGCCCGCGACGACGACGACAAGACGACGCCAGAGCGCAC 2012  
Db 2741 TTCGATCAGCTGGGCGAAGCCCGCGACGACGACGACAAGACGACGCCAGAGCGCAC 2800  
QY 2013 GGAAGACGTAGCGGCGTAG 2031  
Db 2801 GGAAGACGTAGCGGCGTAG 2819  
  
RESULT 6  
PCL414670 5715 bp DNA circular SYN 19-DEC-2001  
LOCUS  
DEFINITION Cloning vector pSET152.  
ACCESSION AJ414670  
VERSION AJ414670.1 GI:17974209  
KEYWORDS aac(3)IV gene; apramycin resistance gene; int gene; integrase; lacZ alpha peptide; lacZ gene.  
SOURCE Cloning vector pSET152.  
ORGANISM Cloning vector pSET152  
REFERENCE 1  
AUTHORS Wilkinson,C.J., Hughes-Thomas,Z.A., Martin (nee Rowe),C.J., Bohm,I., Mironenko,T., Deacon,M., Wheatcroft,M., Wirtz,G., Staunton,J. and Leadlay,P.F.  
TITLE Increasing the efficiency of heterologous promoters in actinomycetes  
JOURNAL J. Mol. Microbiol. Biotechnol.  
REFERENCE 2 (bases 1 to 5715)  
AUTHORS Wilkinson,C.J.  
TITLE Direct Submission  
JOURNAL Submitted (28-SEP-2001) Wilkinson C.J., Department of Biochemistry, University of Cambridge, 80 Tennis Court Rd., Cambridge, CB2 1GA, UNITED KINGDOM  
COMMENT Related sequences V00635, J02451, M24524, M21778.  
FEATURES  
source location/Qualifiers  
1..5715  
/organism="Cloning vector pSET152"  
/specific\_host="Streptomyces sp."  
/db\_xref="taxon:173260"  
/lab\_host="Escherichia coli"  
complement(163..507)  
/gene="lacZa"  
complement(163..507)  
/gene="lacZa"  
/function="blue-white screening"  
/codon\_start=1  
/transl\_table=11  
/product="lacZ alpha peptide"  
/protein\_id="CAC93946.1"  
/db\_xref="GI:17974210"  
/translation="MITNISRAADPLESTCSPLALAVLQRDWENPGVTQLNRLAAHPFASWRNSEEARTRDPSQQLRSLNGEWMRLMYFLTLGCSHRINSPMSSTSGIGSAADAKCRST"  
complement(508..509)  
/note="pUC18"  
/replace="ggt"  
  
variation

```
gene      1729..2505
          /gene="aac(3)IV"
CDS       1729..2505
          /gene="aac(3)IV"
          /function="apramycin resistance"
          /codon_start=1
          /transl_table=11
          /product="apramycin resistance gene"
          /protein_id="CAC93947.1"
          /db_xref="GI:17974211"
          /translation="MQYEWKAEIIGQLNLGVTPGVLVHSSFRSVRPLEDPLGL
          IEALRAALPGGTLVMPMSWGLDEPEDPATSPVTPDLGVSDTFWRLPNVKSAPHE
          AFAAAGPOAEQIISDPLPLPHSPASPVARVHLDGOVLLGVGHDAANTLHLAEI
          KVPYGVPRHCTILODGKLVRYDYLENDHCERFALDRMLKEKSLQKEGPGVGHAFARL
          IRSRDIYATALGQLGRDPLIFLHPPEAGCEECDAARQSIG"
rep_origin 3340..3451
          /note="RP4 origin of single-stranded DNA transfer"
          /direction=RIGHT
          /translation="RIGHT"
          /db_xref="GI:17974211"
          /note="phic31 attP site"
          /gene="phic31 attP site"
          /gene="int"
          /gene="int"
          /gene="int"
          /function="phic31 integrase"
          /codon_start=1
          /transl_table=11
          /product="integrase"
          /protein_id="CAC93948.1"
          /db_xref="GI:17974212"
          /translation="MTQGVTVGVDTYAGAYDROSRERENSSAASPATORSANEDKAD
          LOREYERDGRFRFVGHSEAPGTSAGTAERPEFERILNECRAGLNMTIVYDVSRE
          SRLKYMDAIPVSESLALGVTVSTQEGVFRQGNVMDLHLMLRDLASHKESLSKSAK
          ILDTKNLQRELGGVYGKAPYGFELVSETEITRNGRMVNLINMLASHSTPLTGPE
          FEEDYIRMMWREIKTHKHLPEKPGSOAIIHPSITGLCKRMDADAVPTRGETTIGKTA
          SSAMDPAVTMRLRDPRAGFAAEVIYKKKPDGTPPTKIEGRIQRPDITLRPELDC
          GPIEPAEWYELQAWLDGRGRGKLSRGQATISAMDKLYCEGAVMTSKGEESIKDS
          YRCRRRKVVDPSPAGQHEGTCNVSMALDKFVAERIEFNKIRHAEDEBETLALLWEAR
          RFGKLTAEPEKSEGERANLVAERADALNALLELYEDRAGAYDGPGRKHFRKQAAALT
          LRQGAERLAELEAEAPKLPIDOWFPEADADPTGPKSWMGRASVDDKRVFVGLFV
          DRIYVTKSTGRGQGFIEKRASITWAKPPTDDEDDAQDGTEDVAA"
```

BASE COUNT 1189 a 1643 c 1745 g 1138 t  
ORIGIN

Query Match 89.2%; Score 1812.6; DB 12; Length 5715;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1815; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 213 GATGTGACAGTACGGCGGTGCTTACAGCCGTGAGTCGCGGAGCGGAGAAATTCAGCGC 272
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3877 GGTGACACGTACGGCGGTGCTTACAGCCGTGAGTCGCGGAGCGGAGAAATTCAGCGC 3936

QY 273 AGCAAGCCCGACGACACAGCGGTACGCCCAAGAGACAAGCGCGCCGACCTTCAGCGCGA 332
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3937 AGCAAGCCCGACGACACAGCGGTACGCCCAAGAGACAAGCGCGCCGACCTTCAGCGCGA 3996

QY 333 AGTCGAGCGGACGCGGGCGGTTCAGGTTGTCGGGCATTTTCAGCGAAGCGCGCGGCAC 392
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3997 AGTCGAGCGGACGCGGGCGGTTCAGGTTGTCGGGCATTTTCAGCGAAGCGCGCGGCAC 4056

QY 393 GTCGGCGTTCCGGAGCGGCGGAGCGCCCGGAGTTCGAACGCATCCTGAAACGATGCCGCGC 452
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4057 GTCGGCGTTCCGGAGCGGCGGAGCGCCCGGAGTTCGAACGCATCCTGAAACGATGCCGCGC 4116

QY 453 CGGGCGGCTCAACATGATCATTTCTATGACGTGTCGGCTTCTCGCGCTGGAAGGTCAAT 512
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4117 CGGGCGGCTCAACATGATCATTTCTATGACGTGTCGGCTTCTCGCGCTGGAAGGTCAAT 4176

QY 513 GGACGCGATTCGATTTGTCGGAATGCTCGCCCTGGGCGTGACGATTTTCCACTCA 572
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4177 GGACGCGATTCGATTTGTCGGAATGCTCGCCCTGGGCGTGACGATTTTCCACTCA 4236

QY 573 GGAAGCGCTCTTCGGGACGGAAGCGTCATGACCTGATTCACCTGATATGCGGCTCA 632
```

```
Db 4237 GGAAGCGCTCTTCGGGACGGAAGCGTCAATGACCTGATTCACCTGATATGCGGCTCA 4296

QY 633 CGCGTCGCACAAAGATCTTCGCTGAAGTCGCGGAAGATTTCTGCACAGCAAGACCTTCA 692
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4297 CGCGTCGCACAAAGATCTTCGCTGAAGTCGCGGAAGATTTCTGCACAGCAAGACCTTCA 4356

QY 693 GCGCGAATTGGCGGGGTACGTCGGCGGGAAGCGCCTTACGCGCTTCGAGCTTGTTCGGA 752
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4357 GCGCGAATTGGCGGGGTACGTCGGCGGGAAGCGCCTTACGCGCTTCGAGCTTGTTCGGA 4416

QY 753 GACGAAGGAGATCACGCGCAACGCGCGGAATGTCATATGTCATCAACAAGCTTGCACA 812
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4417 GACGAAGGAGATCACGCGCAACGCGCGGAATGTCATATGTCATCAACAAGCTTGCACA 4476

QY 813 CTCGACCACTCCCTTACCGGACCTTCGAGTTTCGAGCCCGAGCTAATCCGTTGTGTTG 872
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4477 CTCGACCACTCCCTTACCGGACCTTCGAGTTTCGAGCCCGAGCTAATCCGTTGTGTTG 4536

QY 873 GCGTGAGATCAAGACGACACAACACCTTCCCTTCAGCCCGGCGAGTCAAGCCCGCATTC 932
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4537 GCGTGAGATCAAGACGACACAACACCTTCCCTTCAGCCCGGCGAGTCAAGCCCGCATTC 4596

QY 933 CCGCGGACGATCACGGGGCTTTGTAAAGCCGATGAGCGCTGACGCCGTGCCGACCCGGGG 992
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4597 CCGCGGACGATCACGGGGCTTTGTAAAGCCGATGAGCGCTGACGCCGTGCCGACCCGGGG 4656

QY 993 CGAGACGATTTGGGAAGAAGACCCGCTTCAAGCGCCTGGGACCCCGGCAACCGTTATGCGAAT 1052
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4657 CGAGACGATTTGGGAAGAAGACCCGCTTCAAGCGCCTGGGACCCCGGCAACCGTTATGCGAAT 4716

QY 1053 CCTTCGGGACCCCGCTATTCGGGGCTTCGCGCGTGAAGTATCTACAAGAAGACCCGGA 1112
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4717 CCTTCGGGACCCCGCTATTCGGGGCTTCGCGCGTGAAGTATCTACAAGAAGACCCGGA 4776

QY 1113 CGGACGCGCGACCCAGCAAGATTTAGGGTTACCGCATTCAGCGGACCCGATCAGCTCCG 1172
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4777 CGGACGCGCGACCCAGCAAGATTTAGGGTTACCGCATTCAGCGGACCCGATCAGCTCCG 4836

QY 1173 GCGGTCGACGCTGATTTGCGGACCGATCATCGAGCCCGCTGAAGTATGACTTCAGGC 1232
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4837 GCGGTCGACGCTGATTTGCGGACCGATCATCGAGCCCGCTGAAGTATGACTTCAGGC 4896

QY 1233 GTGTTGACGCGGACGCGGGCGCGCAAGGGCTTCCCGGGGCAAGCCATTTCTGTCCGC 1292
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4897 GTGTTGACGCGGACGCGGGCGCGCAAGGGCTTCCCGGGGCAAGCCATTTCTGTCCGC 4956

QY 1293 CATGCAAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAAGCGCGGGGAAGAATC 1352
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4957 CATGCAAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAAGCGCGGGGAAGAATC 5016

QY 1353 GATCAAGGACTTCTACCGCTGCGCTCGCGGGAAGGTGTCGACCCGTCCTCACCTGGGCA 1412
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5017 GATCAAGGACTTCTACCGCTGCGCTCGCGGGAAGGTGTCGACCCGTCCTCACCTGGGCA 5076

QY 1413 GCACGAAGGACGTCGACAGCTACGATGGCGGCACTCGACAAGTTCGTTGCGGAACGAT 1472
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5077 GCACGAAGGACGTCGACAGCTACGATGGCGGCACTCGACAAGTTCGTTGCGGAACGAT 5136

QY 1473 CTTCAACAAGATCAGGACGCGCGGAAGGCGACGAAGAGAGAGCTTGGCGCTTCTGTGGGAAGC 1532
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5137 CTTCAACAAGATCAGGACGCGCGGAAGGCGACGAAGAGAGAGAGCTTGGCGCTTCTGTGGGAAGC 5196

QY 1533 CGCCCGACGCTTCGGCAAGCTCACTGAGGCGCTGAGAAGAGCGCGGAACGCGGAACCT 1592
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5197 CGCCCGACGCTTCGGCAAGCTCACTGAGGCGCTGAGAAGAGCGCGGAACGCGGAACCT 5256

QY 1593 TGTTCGGAGCGCGCGCAAGCCCTGAAGCGCCTTGAAGAGCTGTACGAAGACCGCGCGGC 1652
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5257 TGTTCGGAGCGCGCGCAAGCCCTGAAGCGCCTTGAAGAGCTGTACGAAGACCGCGCGGC 5316

QY 1653 AGCGCGTACGACGAGCCCGTTGGCAGGAAGCACTTCGGAAGCAACAGCGACGCGTGA 1712
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



Db 5317 AGCGCGGTACGACGGACCCGTTGGCAGGAAGCACTTCCGGAAGCAACAGGACCGCTGAC 5376

Oy 1713 GCTCCGGCAGCAAGGGCGGAAGCGGCTTGCCGAACCTGAAGCCGCCGAAGCCCGAA 1772

Db 5377 GCTCCGGCAGCAAGGGCGGAAGCGGCTTGCCGAACCTGAAGCCGCCGAAGCCCGAA 5436

Oy 1773 GCTTCCCTTGACCAATGTTCCCGGAAGCGCGACGCTGACCCGACCGCCCTAAGTC 1832

Db 5437 GCTTCCCTTGACCAATGTTCCCGGAAGCGCGACGCTGACCCGACCGCCCTAAGTC 5496

Oy 1833 GTGTGGGGGGCGCGCTCAGTAGACGACAGCGCGTGTTCGTCGGGCTCTTCGTAGACAA 1892

Db 5497 GTGTGGGGGGCGCGCTCAGTAGACGACAGCGCGTGTTCGTCGGGCTCTTCGTAGACAA 5556

Oy 1893 GATCGTGTTCACGAAGTCGACTACGGGCGAGGGGCGAGGGAACGCCCATCGAGAGCGCGC 1952

Db 5557 GATCGTGTTCACGAAGTCGACTACGGGCGAGGGGCGAGGGAACGCCCATCGAGAGCGCGC 5616

Oy 1953 TTGCATCAGCTGGGGCGAAGCCCGCCAGCAGCAGCAGCAGCAGCGCCAGAGCGGCAC 2012

Db 5617 TTGCATCAGCTGGGGCGAAGCCCGCCAGCAGCAGCAGCAGCAGCGCCAGAGCGGCAC 5676

Oy 2013 GGAAGACGTAGCGGCGTAG 2031

Db 5677 GGAAGACGTAGCGGCGTAG 5695

RESULT 7

LOCUS BPH6589 41489 bp DNA linear PHG 12-MAY-2000

DEFINITION Bacteriophage phi-C31 complete genome.

ACCESSION AJ006589

VERSION AJ006589.1 GI:3947449

KEYWORDS C gene; dCMP deaminase; deoxynucleotide monophosphate kinase; DNA polymerase; int gene; integrase; large subunit; major capsid protein; major tail protein; portal protein; protease; putative; repressor; tail fibre protein; tail tape measure protein; terminase; transfer RNA-Thr; tRNA-Thr.

SOURCE Bacteriophage phi-C31.

ORGANISM Bacteriophage phi-C31.

Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae; Lambda-like viruses.

REFERENCE 1 (bases 1 to 41489)

AUTHORS Hendrix,R.W., Smith,M.C., Burns,R.N., Ford,M.E. and Hatfull,G.F.

TITLE Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage

Proc. Natl. Acad. Sci. U.S.A. 96 (5), 2192-2197 (1999)

JOURNAL MEDLINE 99162580

PUBMED 10051617

REFERENCE 2 (bases 1 to 41489)

AUTHORS Smith,M.C., Burns,R.N., Wilson,S.E. and Gregory,M.A.

TITLE The complete genome sequence of the Streptomyces temperate phage straight phiC31: evolutionary relationships to other viruses

Nucleic Acids Res. 27 (10), 2145-2155 (1999)

JOURNAL MEDLINE 99238410

PUBMED 10219087

REFERENCE 3 (bases 1 to 41489)

AUTHORS Smith,M.C.M.

TITLE Direct Submission

JOURNAL Submitted (01-JUN-1998) Smith M.C.M., Genetics, University of Nottingham, Queens Medical Centre, Nottingham, NG7 2UH, UK

FEATURES location/Qualifiers

source

1. 41489

/organism="Bacteriophage phi-C31"

/virion

/strain="Norwich stock"

/db\_xref="taxon:10719"

gene

CDS

/gene="31"

145..585

/gene="31"

/codon\_start=1

/transl\_table=11

/product="gp31"

/protein\_id="CAA07101.1"

/db\_xref="GI:3947450"

/db\_xref="SPTREMBL:Q37836"

/translation="MSRAKSPELRTGNANAAESAAPVVEGRAPRVPAHLKATGKV WRNVQAGMGAYSPDTRNVILRYCELHDRADLISLEADGYMSEGINQOPVAHPML RYVSTEKELRSIETAIETPEARMRLGIVAEARKVAAGPEDF"

complement(640..1170)

/gene="32"

complement(640..1170)

/gene="32"

/note="transcribed in lysogens"

/codon\_start=1

/transl\_table=11

/product="gp32"

/protein\_id="CAA07102.1"

/db\_xref="GI:3947451"

/db\_xref="SPTREMBL:Q37837"

/translation="MNAATPTRYFGKTNGLHYVRVLRGEELRYTPGYTLDMAVD AACWEAFRVVAPAEPTPAEKRDMMKRDGAAGVDKPEGLREGDYVWHTYTMKAAYNRKG QRVSEYQVSEWEVSCGAWARDWARGLYGYRLVGFTEGAYANKGETVYVHRDPKALKR AEAETRAVYVAAFASKR"

CDS

gene

CDS

1282..2841

/gene="33"

1282..2841

/gene="33"

/function="putative large terminase subunit"

/codon\_start=1

/transl\_table=11

/product="gp33"

/protein\_id="CAA07103.1"

/db\_xref="GI:3947452"

/translation="MADWAGIDPVIARHIPADATVPSEGYRVAKWIEECYLTGSFAG OPEFLPWORELIDAVVLTODTEGRWRKRRVYVVCVARKNGKSTIAAIMLYLIA DRGDAQROIIAANDRNQARMVFDKAKQMVNASPKLAACDVQORDVIRKNDTRYVVS ADAGROGGLNPAAVSLDEYAFSKHSDLFDALTLGSAARNQPMFLISTAGPDPDGPEA ALCEQGERVNSGEADPTLYRSGWPKLGETVDHDPVWRACNPSYDLINPDDFKAA AQRSTEASFRIYRLSQFVFGASTWLPBGLMDSLADDDPLEGDEVYLGFDGSKGDS TALVACRIRDLKVFVLGWEAPADDAHWRVPMADVAREELHTALDVRVRLVADPYRW EETIDNLEADGFPVEAFPTNSLARMVPATQAVYDACRDGRLSHDGNPALGRHIGNAVL KEDARGARITKEHAASSRRKIDLAVAMVLAVHGAVMWRDNGIVSDKPIATWEDDEGN VEVHPDHAEEF"

gene

CDS

2886..4259

/gene="34"

2886..4259

/gene="34"

/function="putative portal protein"

/codon\_start=1

/transl\_table=11

/product="gp34"

/protein\_id="CAA07104.1"

/db\_xref="GI:3947453"

/db\_xref="SPTREMBL:Q92XB2"

/translation="MGFMSALFGRGHSPALDGIETARAWEPYDPSIYNLGAVASGETV TPDALQVSAVFASVRLSETIATLPLSTYSKRGSRKEIVTPEWLDYPAEPPGGMGR IDILISQTVLSLLQGNATLAVRWQGNIVGLVDPTKIHVHMVMVDGLRRKVEAYD IDADGNEVLLGMFTPRDYLHIPGMLPGDFVGCSPISYARESISGLAOKYGSKFPA NGAMPGAVEVEPGTMSSEGLARAREAWRAANGVDNAHRAVALLTEGAFSKVAMSPDE AOFLOTROFOVPEIARIPIGVPHLISDATNSTWSGLAEONIAFTMFSLRPWLRIE AGFNRLLEAETADRFERYKFNLDIEIKRGAPKERMELMSGLONGIYSIDEVRAEDMT PLPDGLGEKRYRVPLNLGEVGEPEPEPAPAPAEPEDEPEPECKPDDEGATE EDDDDDA"

gene

CDS

4240..4881

/gene="35"

4240..4881

/gene="35"

/note="putative protease"

/codon\_start=1

/transl\_table=11

/product="gp35"

/protein\_id="CAA07105.1"

/db\_xref="GI:3947454"

/db\_xref="SPTREMBL:Q92XB1"

```
gene
CDS
      /translation="MTKTRRLRAVAVTLERASDDGRISMRCYAYRFNELSHDLGGE
      RERIVPGAGAPSLRONDVYATFNHTRSLLGRTSSGTLRVEDREGWYEIDLPDITV
      GRDVAKLLKRGDLQSSFTFRVLGGQRADDDDPETGLPREITAMDVVELGYPVNP
      AYPPTQASLRISIEETLRIGEFAPTEERDSQPDGDVAPASHFDARSLVRALSK"
      4893..6071
      /gene="36"
      4893..6071
      /gene="36"
      /function="major capsid protein"
      /codon_start=1
      /transl_table=11
      /product="gp36"
      /protein_id="CAA07106.1"
      /db_xref="GI:3947455"
      /db_xref="SPTREMBL:Q9ZNV6"
      /translation="MDATTLSEANFEARERATAELRSLTDEFAGKEMTAAREKEERLL
      TAVADFDGRIRKIDAIKATDAVTSLLSLGQSSGSAORSADHDDAVLRAGNLGEAR
      SEEFAPERDGTAGNPNVLSRFLYGLIAQAVERSAIMRGASFTTSDANPMDFTV
      ITGRATAGIVGETAEIPESYPATTQSRMGFKYGFASVSYEFATDQVLIDLGLFVSD
      AGPAIGDAGRHFLLTGTGTPRGILTDATGANAFGEADADSKVSDALIDLFHEVPS
      AYRKNKRVVNDLRAAQMRKLKDANGQYLMQSALTYGAPDTFNGKVVEITDDGMADKV
      LFADLSKYRVRFAGSLRVDRSDAKFSTDQIVYRFLQRADGLLVDAARGAKYLVTPPA"
      5226..6068
      /gene="36"
      /product="gp36"
      6144..6512
      /gene="37"
      6144..6512
      /gene="37"
      /codon_start=1
      /transl_table=11
      /product="gp37"
      /protein_id="CAA07107.1"
      /db_xref="GI:3947456"
      /db_xref="SPTREMBL:Q9ZXBO"
      /translation="MAYATIEELRALDGLDSDALFSEDLISDAIDFSYTEVEAYCGRK
      WDTAEDPTPETIRWCVRITLAROYVLHDVSRIPDRALQLOSEFGSIQLAQAGGNWRPTS
      *LPEVNAKLNLYRVRLPFIFM"
      6519..6956
      /gene="38"
      6519..6956
      /gene="38"
      /codon_start=1
      /transl_table=11
      /product="gp38"
      /protein_id="CAA07108.1"
      /db_xref="GI:3947457"
      /db_xref="SPTREMBL:Q9ZXAX9"
      /translation="MALIFNAKVRLEALKANYPGDVOCFAETGDNRRKQVWLIGAT
      VDDLAPVAMRSGAKPTNVTGYEHAHAVTTPGNPVDERAERVGIRDIYVKAACAAYNA
      DLVSVPLMDVRPESASVESTETTDGAYSALTVRVVRGRVYQ"
      6989..7975
      /gene="39"
      6989..7975
      /gene="39"
      /function="major tail protein"
      /codon_start=1
      /transl_table=11
      /product="gp39"
      /protein_id="CAA07109.1"
      /db_xref="GI:3947458"
      /db_xref="SPTREMBL:Q9ZXAX8"
      /translation="MALDASIGIGREDITYGLSGVEVEGEGADSWKQTEPIESVGF
      RAGQOTARADRRNMVNMGGELIDIVLDAGAGSLTAAFDKVTVDTGKVTTLVET
      SDVSEAPSTAQMYRPGVDGSKVAFKHLGCVAATEWSLTAVEBAVKLAVTFDRDATH
      TSTPAEIVEPTYPAEAYPYDWTRTAVELSRAGSVVEFDATSELTGELGKMDRRLN
      GSELKKPVRNALPTYEGLTEGEFSASLGLYDAFVSGEVCSEKATFFGILPGSSLSV
      EAPAIQFTGESPEAATDEVTIHNLPFRVLDPGDGVTPAIKLTIVEPPTPEP"
      7975..8406
      /gene="40"
      7975..8406
      /gene="40"
      /gene="40"
mat_peptide
      5226..6068
      /gene="36"
      /product="gp36"
      6144..6512
      /gene="37"
      6144..6512
      /gene="37"
      /codon_start=1
      /transl_table=11
      /product="gp37"
      /protein_id="CAA07107.1"
      /db_xref="GI:3947456"
      /db_xref="SPTREMBL:Q9ZXBO"
      /translation="MAYATIEELRALDGLDSDALFSEDLISDAIDFSYTEVEAYCGRK
      WDTAEDPTPETIRWCVRITLAROYVLHDVSRIPDRALQLOSEFGSIQLAQAGGNWRPTS
      *LPEVNAKLNLYRVRLPFIFM"
      6519..6956
      /gene="38"
      6519..6956
      /gene="38"
      /codon_start=1
      /transl_table=11
      /product="gp38"
      /protein_id="CAA07108.1"
      /db_xref="GI:3947457"
      /db_xref="SPTREMBL:Q9ZXAX9"
      /translation="MALIFNAKVRLEALKANYPGDVOCFAETGDNRRKQVWLIGAT
      VDDLAPVAMRSGAKPTNVTGYEHAHAVTTPGNPVDERAERVGIRDIYVKAACAAYNA
      DLVSVPLMDVRPESASVESTETTDGAYSALTVRVVRGRVYQ"
      6989..7975
      /gene="39"
      6989..7975
      /gene="39"
      /function="major tail protein"
      /codon_start=1
      /transl_table=11
      /product="gp39"
      /protein_id="CAA07109.1"
      /db_xref="GI:3947458"
      /db_xref="SPTREMBL:Q9ZXAX8"
      /translation="MALDASIGIGREDITYGLSGVEVEGEGADSWKQTEPIESVGF
      RAGQOTARADRRNMVNMGGELIDIVLDAGAGSLTAAFDKVTVDTGKVTTLVET
      SDVSEAPSTAQMYRPGVDGSKVAFKHLGCVAATEWSLTAVEBAVKLAVTFDRDATH
      TSTPAEIVEPTYPAEAYPYDWTRTAVELSRAGSVVEFDATSELTGELGKMDRRLN
      GSELKKPVRNALPTYEGLTEGEFSASLGLYDAFVSGEVCSEKATFFGILPGSSLSV
      EAPAIQFTGESPEAATDEVTIHNLPFRVLDPGDGVTPAIKLTIVEPPTPEP"
      7975..8406
      /gene="40"
      7975..8406
      /gene="40"
      /gene="40"
```

Query Match									
Best Local Similarity 89.2%; Score 1812.6; DB 7; Length 41489;									
Matches 1815; Conservative 99.8%; Pred. No. 0;									
Mismatch 0; Mismatches 4; Indels 0; Gaps 0;									
gene									
8406..8878									
/codon_start=1									
/transl_table=11									
/product="gp40"									
/protein_id="CAA07110.1"									
/db_xref="GI:3947459"									
/db_xref="SPTREMBL:Q92XA7"									
/translation="MAQRSAYTIQVDLROFORNVRLRDKELNKAVREANKASGEVL									
IPOAKHESPDGHRDPKSSKRYRPGKLDKSIKVTASAKAVIKAGSARVPYAAAIHFG									
YRKNISANRFLYRAMARKSDVVAATYERRIAAVEKYLE"									
8406..8878									
QY	213	GATGTCAGGTACGGGGGTGCTTACGACCGTCAGTCGCGCGGAGCGGAGAAATTCGAGCGC	272						
Db	38444	GGTGACACAGTACGGGGGTGCTTACGACCGTCAGTCGCGCGGAGCGGAGAAATTCGAGCGC	38503						
QY	273	AGCAAGCCAGCGACACAGCGTAGCGCCCAAGAGACAGAGCGGCGGACCTTCAGCGCGA	332						
Db	38504	AGCAAGCCAGCGACACAGCGTAGCGCCCAAGAGACAGAGCGGCGGACCTTCAGCGCGA	38563						
QY	333	AGTCGAGCGCGAGCGGGGGCGGTTGCTGAGGCTTCGCGGCAATTCAGCGAGCGCGGCGAC	392						
Db	38564	AGTCGAGCGCGAGCGGGGGCGGTTGCTGAGGCTTCGCGGCAATTCAGCGAGCGCGGCGAC	38623						
QY	393	GTCGGCGTTGCGGAGCGCGGAGCGCGCGGAGTTGCAACGCATTCGAGCAATGCCGCGC	452						
Db	38624	GTCGGCGTTGCGGAGCGCGGAGCGCGCGGAGTTGCAACGCATTCGAGCAATGCCGCGC	38683						
QY	453	CGGGCGGCTCACATGATCATGTCATGATGAGCGTGTGCGGCTTCGCGCTGAGAGTCAAT	512						
Db	38684	CGGGCGGCTCACATGATCATGTCATGATGAGCGTGTGCGGCTTCGCGCTGAGAGTCAAT	38743						
QY	513	GGACGCGATTCGATTTGCTCGGAATGCTCGCCCTGGGCGTGACGATTGTTCCACTCA	572						
Db	38744	GGACGCGATTCGATTTGCTCGGAATGCTCGCCCTGGGCGTGACGATTGTTCCACTCA	38803						
QY	573	GGAAGCGCTTCGCGGAGGGAACGTATGACCTGATTCACCTGATTATGCGGCTCGA	632						
Db	38804	GGAAGCGCTTCGCGGAGGGAACGTATGACCTGATTCACCTGATTATGCGGCTCGA	38863						
QY	633	CGGTCGACAAAGAATCTTCGTCGAAGTCGGGAGATTCGACACGAGAACCTTCA	692						
Db	38864	CGGTCGACAAAGAATCTTCGTCGAAGTCGGGAGATTCGACACGAGAACCTTCA	38923						
QY	693	GCGGCAATTTGGGGGCTACGTCGGCGGGAAGCGGCTTACGGCTTCGAGCTTGTTCCGA	752						
Db	38924	GCGGCAATTTGGGGGCTACGTCGGCGGGAAGCGGCTTACGGCTTCGAGCTTGTTCCGA	38983						
QY	753	GACGAGGAGATCACGGCGCAACGGCCGAATGTCATATGTCATCAACAACCTTGGCA	812						
Db	38984	GACGAGGAGATCACGGCGCAACGGCCGAATGTCATATGTCATCAACAACCTTGGCA	39043						
QY	813	CTGACCACTCCCTTACCGGACCTTCGAGTTGAGCCCGACGTAATCCGCTGTGTG	872						
Db	39044	CTGACCACTCCCTTACCGGACCTTCGAGTTGAGCCCGACGTAATCCGCTGTGTG	39103						
QY	873	GCGTGATCAAGAGCGACAAACCTTCCCTTCAAGCCGGGAGTCAAGCCGCATTCA	932						
Db	39104	GCGTGATCAAGAGCGACAAACCTTCCCTTCAAGCCGGGAGTCAAGCCGCATTCA	39163						
QY	933	CCCGGAGCATCACGGGGCTTGTAAAGCGCATGACGCTGACGCCGTGCCGACCCGGG	992						
Db	39164	CCCGGAGCATCACGGGGCTTGTAAAGCGCATGACGCTGACGCCGTGCCGACCCGGG	39223						
QY	993	CGAGACGATTGGAGAGAACCCGCTTCAAGCGCTGGGACCCCGCAACCCGTTATGCGAAT	1052						
Db	39224	CGAGACGATTGGAGAGAACCCGCTTCAAGCGCTGGGACCCCGCAACCCGTTATGCGAAT	39283						
QY	1053	CCITCGGAGACCCGCGTATTCGGGGCTTCGCGCTGAGGTGATCTACAAGAGACCCGGA	1112						

Db 39284 CCTTCGGGACCCCGCTATTGCGGGCTTCGCCGCTGAGGTGATCTACAGAGAAGCCGGA 39343  
QY 1113 CGGCACGCCGACACGAGATTGAGGGTTACCGCATTCAGCGGACCCGATCACGCTCCG 1172  
Db 39344 CGGCACGCCGACACGAGATTGAGGGTTACCGCATTCAGCGGACCCGATCACGCTCCG 39403  
QY 1173 GCCGCTGAGCTTGATTGCGGACCGATCATCGAGCCCGCTGAGTGTATGAGCTTCAGGC 1232  
Db 39404 GCCGCTGAGCTTGATTGCGGACCGATCATCGAGCCCGCTGAGTGTATGAGCTTCAGGC 39463  
QY 1233 GTGCTTGACGGCAGGGGGGGCGGCAAGGGGCTTCGCCGGGCAAGCCATTCTGTCGC 1292  
Db 39464 GTGCTTGACGGCAGGGGGGGCGGCAAGGGGCTTCGCCGGGCAAGCCATTCTGTCGC 39523  
QY 1293 CATGACACAAGCTGTACTGCGAGTGTGGCGCCGTCATGACTTCGAAGCGCGGGAAGAATC 1352  
Db 39524 CATGACACAAGCTGTACTGCGAGTGTGGCGCCGTCATGACTTCGAAGCGCGGGAAGAATC 39583  
QY 1353 GATCAAGGACTCTTACCGCTGCGCTGCGGGAAGGTGTCGACCCGTCGCACTGGGCA 1412  
Db 39584 GATCAAGGACTCTTACCGCTGCGCTGCGGGAAGGTGTCGACCCGTCGCACTGGGCA 39643  
QY 1413 GCACGAAGGACGTCGACAGTCAGCATGGCGGCACTGACAAGTTCTGCGGAACGCAT 1472  
Db 39644 GCACGAAGGACGTCGACAGTCAGCATGGCGGCACTGACAAGTTCTGCGGAACGCAT 39703  
QY 1473 CTTCACAAGATCAGGACCGCCGAAGGCGACGAAGACGTTGGCGCTTCTGTGGGAAGC 1532  
Db 39704 CTTCACAAGATCAGGACCGCCGAAGGCGACGAAGACGTTGGCGCTTCTGTGGGAAGC 39763  
QY 1533 CGCCCGACGCTTCGGCAAGCTCACTGAGGCGCTGAGAAGAGCGGGCGAAGCGCGAACCT 1592  
Db 39764 CGCCCGACGCTTCGGCAAGCTCACTGAGGCGCTGAGAAGAGCGGGCGAAGCGCGAACCT 39823  
QY 1593 TGTTCGGGAGCGCGCGCGCCCTGAACGCCCTGAAGAGCTGTACGAAGCGCGCGGC 1652  
Db 39824 TGTTCGGGAGCGCGCGCGCCCTGAACGCCCTGAAGAGCTGTACGAAGCGCGCGGC 39883  
QY 1653 AGCGCGTACGACGAGACCCGTTGGCAGGAAGACACTTCGGAAGCAACAGGACGCGTGAC 1712  
Db 39884 AGCGCGTACGACGAGACCCGTTGGCAGGAAGACACTTCGGAAGCAACAGGACGCGTGAC 39943  
QY 1713 GCTCCGGCAGCAAGGGGGCGGAAGCGGCTTGCCGAACCTTGAAGCGCGCAAGCCCGAA 1772  
Db 39944 GCTCCGGCAGCAAGGGGGCGGAAGCGGCTTGCCGAACCTTGAAGCGCGCAAGCCCGAA 40003  
QY 1773 GCTTCCTTGAACCATGTTCCCGGAAGACGCGCGCTGACCCGACCGCCCTAAGTC 1832  
Db 40004 GCTTCCTTGAACCATGTTCCCGGAAGACGCGCGCTGACCCGACCGCCCTAAGTC 40063  
QY 1833 GTGCTGGGGGGCGCGCTCAGTAGACGACAGCGCGCTGCTCGGGCTCTCTAGACAA 1892  
Db 40064 GTGCTGGGGGGCGCGCTCAGTAGACGACAGCGCGCTGCTCGGGCTCTCTAGACAA 40123  
QY 1893 GATCGTTGTCAGGAAGTCAGTACGGGCGAGGGGCGAGGGAACGCCCATCGAGAAGCGCGC 1952  
Db 40124 GATCGTTGTCAGGAAGTCAGTACGGGCGAGGGGCGAGGGAACGCCCATCGAGAAGCGCGC 40183  
QY 1953 TTTCGATCAGCTGGGCGAAGCCCGCAGCAGCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2012  
Db 40184 TTTCGATCAGCTGGGCGAAGCCCGCAGCAGCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 40243  
QY 2013 GGAAGACGTAGCGGCTAG 2031  
Db 40244 GGAAGACGTAGCGGCTAG 40262

RESULT 8  
LOCUS EVU84006 5967 bp DNA circular SYN 03-MAY-1997  
DEFINITION Expression vector pBSII-LUCINT firefly luciferase (LUCINT),  
beta-galactosidase (lacZ) and beta-lactamase (ampR) genes, complete  
cds and lac operon promoter sequence.

ACCESSION U84006  
VERSION U84006.1 GI:2071944  
KEYWORDS  
SOURCE Expression vector pBSII-LUCINT.  
ORGANISM Expression vector pBSII-LUCINT.  
REFERENCE 1 (bases 1 to 5967)  
AUTHORS Mankin,S.L., Allen,G.C. and Thompson,W.F.  
TITLE Introduction of a Plant Intron into the Luciferase Gene of Photinus  
pyralis  
JOURNAL Plant Mol. Biol. Rep. (1997) In press  
REFERENCE 2 (bases 1 to 5967)  
AUTHORS Mankin,S.L., Allen,G.C. and Thompson,W.F.  
TITLE Direct Submission  
JOURNAL Submitted (06-JAN-1997) Botany, NC State Univ., Box 7612, Raleigh,  
NC 27695-7612, USA  
FEATURES  
source  
1. 5967  
/organism="Expression vector pBSII-LUCINT"  
/db\_xref="taxon:56405"  
/plasmid="pluk07"  
/note="chimera plasmid consisting of a pBS-type plasmid  
(pluk07, dsDNA) derived from the LUC\* gene (a synthetic  
cDNA gene) by the addition of a plant intron sequence"  
52..851  
/note="Camv 35S promoter"  
875..3052  
/gene="LUCINT"  
join(875..1029,1219..2716)  
/gene="LUCINT"  
/EC\_number="1.13.12.7"  
/note="luciferin 4-monooxygenase; LUC\* with PIV2 intron"  
/codon\_start=1  
/product="firefly luciferase"  
/protein\_id="AAB53627.1"  
/db\_xref="GI:2071945"  
/translation="MEDAKNIKKGPAPYPLEDGTAGEOLHKMKRYALVPGTIAFTD  
AHIEVNITYAEYFEMSRYLAEMKRYGLTNHRIVVCSENSLOFFMPVLGALFIGVAV  
APANDIYNERELNSMNIISOPTVVFVSKGLQKILNVOKLPILOKIIIMDSKTDYOG  
FQSMYTFVTSHPGNEFYDEVPESFDRDKTIALIMNSGSGTLPKVALPHRTACVR  
FSHARDPIFGNQLIPDAILSVVPHFGMETTIGYLICGRVLMYRFEELFLRS  
LQDYKIQSALVPTLFSFPAKSTLIDKYDLSNLHEIASGAPLSKEVGEAVAKRHLPL  
GIRQGYLTETSAIILITPEGDDKPGAVKVPVPEFAKVVDLTGKTLGVNORGEICV  
RGPIMISGVNNEPATNALIDKDGWLHSGDIAYWDEDEHFIYDRLSLIKYGIOYA  
PALESILQHPNIFDAGVAGLPDDAGELPAVAVVLEHGMTKEIIVDVASQVTT  
AKLRGVVVFVEVPKGLTGKLDARKIREILIKAKKGKSKL"  
1030..1218  
/gene="LUCINT"  
/note="PIV2 chimeric intron"  
2786..3052  
/gene="LUCINT"  
/standard\_name="nopoline synthase 3' region"  
/note="nos"  
2816..2821  
/gene="LUCINT"  
3146..3355  
/gene="lacZ"  
3146..3355  
/gene="lacZ"  
/codon\_start=1  
/transl\_table=11  
/product="beta-galactosidase"  
/protein\_id="AAB53629.1"  
/db\_xref="GI:2071947"  
/translation="MAVVLRDWDENPGVTQLNRLAHPFASWRNSEARTDRPSOO  
LRSINGEWREIVNVNILLKALNFC"  
3895..4755  
/gene="ampR"  
3895..4755  
/gene="ampR"  
/function="ampicillin resistance marker"  
/codon\_start=1  
/transl\_table=11





QY 150 CTGTAGTTTAAAGTGTGATATTTTAAATTAACTTTTCTAATATATGACCAAAATTT 209  
|||||  
Db 503 CTGTAGTTTAAAGTGTGATATTTTAAATTAACTTTTCTAATATATGACCAAAATTT 562  
QY 210 GTTGATGTGCAGTA 224  
|||||  
Db 563 GTTGATGTGCAGTA 577

RESULT 15

AR037156

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

AR037156 2633 bp. DNA linear PAT 29-SEP-1999  
Sequence 2 from patent US 5801027.  
AR037156  
AR037156.1 GI:5955012  
Unknown.  
Unknown.  
Unclassified.  
1 (bases 1 to 2633)  
Bennett, M., May, S. and Ramsay, N.  
Method of using transactivation proteins to control gene expression  
in transgenic plants  
Patent: US 5801027-A 2 01-SEP-1998;  
Location/Qualifiers  
1. 2633  
/organism="unknown"  
688 a 618 c 671 g 627 t 29 others

Query Match

Best Local Similarity 100.0%; Pred. No. 3.8e-26;

Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 TACGTAAGTTTCGCTTCTACCTTTGATATATATATAATATATCATTAATTAGTAGTA 89  
|||||  
Db 693 TACGTAAGTTTCGCTTCTACCTTTGATATATATATAATATATCATTAATTAGTAGTA 752  
QY 90 ATATAATATTCAAAATATTTTTCAAAATAAAGAATGTAGTATATAGCAATTGCTTTT 149  
|||||  
Db 753 ATATAATATTCAAAATATTTTTCAAAATAAAGAATGTAGTATATAGCAATTGCTTTT 812  
QY 150 CTGTAGTTTAAAGTGTGATATTTTAAATTAACTTTTCTAATATATGACCAAAATTT 209  
|||||  
Db 813 CTGTAGTTTAAAGTGTGATATTTTAAATTAACTTTTCTAATATATGACCAAAATTT 872  
QY 210 GTTGATGTGCAGTA 224  
|||||  
Db 873 GTTGATGTGCAGTA 887

Search completed: May 11, 2003, 21:25:33  
Job time : 6062.5 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 14:41:33 ; Search time 511.5 Seconds  
(without alignments)  
8941.952 Million cell updates/sec

Title: US-09-940-550A-9  
Perfect score: 2031  
Sequence: 1 atggcacaagggtgtgtgac.....cggaaagacgtagcgcgtag 2031

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2031	100.0	2031	ABK12555	DNA encoding novel
2	2031	100.0	9880	ABK12557	Plasmid pBPS EW051
3	2015	99.2	2031	ABK12556	DNA encoding novel
4	1812.6	89.2	5711	AAD04935	C31-Int expression
5	1811	89.2	1842	AAH74879	Nucleotide sequence
6	1809.4	89.1	3401	AAQ06904	Sequence encoding
7	1806.2	88.9	1842	AAH74880	Nucleotide sequence
8	1801.6	88.7	1839	AAH74881	Nucleotide sequence
9	1783	87.8	1833	AAH74882	Nucleotide sequence

10	198.6	9.8	912	22	AAD19829	Green fluorescent
11	198.6	9.8	1577	22	AAD19834	Promoter-reporter
12	198.6	9.8	1618	22	AAD19832	Promoter-reporter
13	198.6	9.8	2943	22	AAD19840	Ubq3(At)-synGFP-N
14	195	9.6	2001	22	AAD19830	Beta-glucuronidase
15	195	9.6	2001	24	ABK15667	Glutathione-S-tran
16	195	9.6	2725	22	AAD19835	Promoter-reporter
17	195	9.6	2730	22	AAD19833	Promoter-reporter
18	195	9.6	4072	22	AAD19841	Ubq3(At)-GIG-NOS t
19	195	9.6	4341	22	AAD19839	Zmubi-GIG-NOS term
20	195	9.6	5534	17	AAT43137	PUMGIT sequence i
21	195	9.6	12817	24	ABK15668	Expression vector
22	195	9.6	13274	24	ABK15666	Binary vector PNOV
23	194	9.6	1701	21	AAZ29120	Plasmid DV127 used
24	194	9.6	3336	21	AAZ29121	Plasmid DV130 comp
25	194	9.6	3877	21	AAZ29123	Plasmid DV132 used
26	192	9.5	1294	22	AAD19907	PTBN comprising Ca
27	190.8	9.4	1691	22	AAD19910	PGI vector based o
28	190	9.4	1034	19	AAV54572	Glucosamine-6-phos
29	181.6	8.9	363	22	AAF32096	Insecticidal prote
30	180.6	8.9	363	22	AAF32093	Insecticidal prote
31	180.6	8.9	363	22	AAF32095	Insecticidal prote
32	180.6	8.9	369	22	AAF32094	Insecticidal prote
33	137.8	6.8	1512	22	AAD19909	PTBN1 vector based
34	137.8	6.8	2708	22	AAD19908	PTBN6 vector based
35	137.4	6.8	1851	22	AAD19912	PGI vector based
36	137.4	6.8	2315	22	AAD19911	Human osteoblast d
37	58.2	2.9	154902	24	ABQ88198	Human osteoblast d
38	53.6	2.6	7752	24	ABL70223	Chemically treated
39	53.6	2.6	159400	24	ABQ88126	Human osteoblast d
40	53	2.6	6641	24	ABL54335	Chemically treated
41	53	2.6	6641	24	ABL32314	Human immune syste
42	52.2	2.6	6118	24	ABN80110	Human chemically m
43	51.8	2.6	1836	24	ABL55639	AMEPV second DNA p
44	51.8	2.6	4654	22	AAS46780	Tumour suppressor
45	51.8	2.6	4654	24	ABL34222	Human immune syste

ALIGNMENTS

RESULT 1	ABK12555	standard; DNA; 2031 BP.
ID	ABK12555	
XX	ABK12555;	
AC	ABK12555;	
XX	18-JUN-2002 (first entry)	
DT		
XX	DNA encoding novel phi C31 integrase, phic31intINT.	
DE		
XX	Phl C31 integrase; recombinase; transgenic; plant;	
KW	agricultural food product; self-excising polynucleotide;	
KW	gene; ds; phic31intINT.	
XX		
OS	Synthetic.	
XX		
PN	WO200216609-A2.	
PD	28-FEB-2002.	
XX		
PF	27-AUG-2001; 2001WO-US26738.	
XX		
PR	25-AUG-2000; 2000US-227961P.	
XX		
PA	(BADI ) BASF PLANT SCI GMBH.	
PA	(MCKE/) MCKERSIE B.	
XX		
PI	Mankin L;	
XX		
DR	WPI; 2002-280939/32.	
XX		
PT	New self-excising polynucleotides, useful for producing transgenic	

Filed at  
1-5-2003  
US

PT plants, removing transgenes from these plants or crops (e.g. food  
PT commodities), and restricting the distribution of transgenes within the  
PT environment  
XX  
PS Claim 5; Fig 6; 60pp; English.  
XX

CC The invention describes an isolated excisable polynucleotide, which  
CC comprises a desired trait polynucleotide and a recombinase polynucleotide  
CC operably linked to a promoter, all flanked by a pair of directly oriented  
CC recombination sites where the recombinase activity is regulatable. The  
CC self-excising polynucleotide is useful for producing transgenic plants,  
CC particularly for removing all transgenic nucleic acid sequences that  
CC encode a gene product from the transgenic plant, thus restoring the  
CC original genetic configuration of the genome. The polynucleotide is also  
CC useful in methods for eliminating unwanted nucleic acids in agricultural  
CC food products and for preventing the escape of certain transgenic traits  
CC into the environment. This sequence encodes the novel phic31 integrase  
CC phic31int, created for use in the self-excising polynucleotide  
CC described in the invention.  
XX  
SQ Sequence 2031 BP; 451 A; 562 C; 640 G; 378 T; 0 other;

Query Match 100.0%; Score 2031; DB 24; Length 2031;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2031; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCACAAGGGGTTGACCGGGGTGATACGTAAGTTCTGCTTCTACCTTGATATA 60  
Db 1 ATGGCACAAGGGGTTGACCGGGGTGATACGTAAGTTCTGCTTCTACCTTGATATA 60  
QY 61 TATATAAT 120  
Db 61 TATATAAT 120  
QY 121 AAAGATGTAGTATATAGCAATTCCTTCTGTAAGTTATAGTGTATATATATATAT 180  
Db 121 AAAGATGTAGTATATAGCAATTCCTTCTGTAAGTTATAGTGTATATATATATATAT 180  
QY 181 ATAATCTTTCTATATATATGACCAAAATTTGTGATGTGACGTAAGGCTGTTACGAC 240  
Db 181 ATAATCTTTCTATATATATGACCAAAATTTGTGATGTGACGTAAGGCTGTTACGAC 240  
QY 241 CGTCAGTCGGCGGAGCGCGAGAAATTCGAGCGCAAGCCGACGACACAGCGTACGCC 300  
Db 241 CGTCAGTCGGCGGAGCGCGAGAAATTCGAGCGCAAGCCGACGACACAGCGTACGCC 300  
QY 301 AACGAAGACAGCGCGCGGACCTTCAGCGGAAGTCGAGCGCGACGCGCGGTTCAAG 360  
Db 301 AACGAAGACAGCGCGCGGACCTTCAGCGGAAGTCGAGCGCGACGCGCGGTTCAAG 360  
QY 361 TTCGTGGGCAATTCAGGAAAGCGCGGCGTTCGCGGACGCGGAGCGCGCGCG 420  
Db 361 TTCGTGGGCAATTCAGGAAAGCGCGGCGTTCGCGGACGCGGAGCGCGCGCGCG 420  
QY 421 GAGTTGCAAGCATCTGAAAGCATGCGCGCGCGGCGCTCAACATGATGATTTGTCTAT 480  
Db 421 GAGTTGCAAGCATCTGAAAGCATGCGCGCGCGGCGCTCAACATGATGATTTGTCTAT 480  
QY 481 GACGTGCGCGCTTCTGCGCGCTGAAGTCAATGACGCGGATTCGATTTGCTCGGAATTG 540  
Db 481 GACGTGCGCGCTTCTGCGCGCTGAAGTCAATGACGCGGATTCGATTTGCTCGGAATTG 540  
QY 541 CTCGCCCTGGGCGTGACGATTTTCCACTCAGGAAGCGCTCTCCGGCAGGGAACGTC 600  
Db 541 CTCGCCCTGGGCGTGACGATTTTCCACTCAGGAAGCGCTCTCCGGCAGGGAACGTC 600  
QY 601 ATGACCTGATTCACCTGATTAATGCGGCTGACGCGCTGCACAAAGATCTTGGCTGAAG 660  
Db 601 ATGACCTGATTCACCTGATTAATGCGGCTGACGCGCTGCACAAAGATCTTGGCTGAAG 660  
QY 661 TCGGCGAAGATTCGACAGCAAGAACCTTCAGCGCGAATTTGGCGGTACGTGCGCGG 720  
Db 661 TCGGCGAAGATTCGACAGCAAGAACCTTCAGCGCGAATTTGGCGGTACGTGCGCGG 720

QY 721 AAGCGCCTTACGGCTTCGAGCTTGTTCGGAGACGAAGAGATTCACGCGCAACGCCGA 780  
Db 721 AAGCGCCTTACGGCTTCGAGCTTGTTCGGAGACGAAGAGATTCACGCGCAACGCCGA 780  
QY 781 ATGTCATGTGTCATTCACAAAGCTTGGCCTACTGACCACTCCCTTACCGGACCTTC 840  
Db 781 ATGTCATGTGTCATTCACAAAGCTTGGCCTACTGACCACTCCCTTACCGGACCTTC 840  
QY 841 GAGTTGAGCCCGACGTAATCCGGTGTGCTGGCTGAGATCAAGACGACAAACACCTT 900  
Db 841 GAGTTGAGCCCGACGTAATCCGGTGTGCTGGCTGAGATCAAGACGACAAACACCTT 900  
QY 901 CCCTTCAAGCGGGCAGTCAAGCCGCTTCAACCCGGGAGCATCACGGGCTTTGTAAG 960  
Db 901 CCCTTCAAGCGGGCAGTCAAGCCGCTTCAACCCGGGAGCATCACGGGCTTTGTAAG 960  
QY 961 CGCATGAGCGCTGACGCGCTGCGGACCCGGGCGAGACGATTGGGAAGAACCGCTTCA 1020  
Db 961 CGCATGAGCGCTGACGCGCTGCGGACCCGGGCGAGACGATTGGGAAGAACCGCTTCA 1020  
QY 1021 AGCGCCTGGGACCCGGCAACCGTTATGCGAATCTTCGGGACCCGCTATTCGGGCTTC 1080  
Db 1021 AGCGCCTGGGACCCGGCAACCGTTATGCGAATCTTCGGGACCCGCTATTCGGGCTTC 1080  
QY 1081 GCGGCTGAGGTGATCTTACAAGAAAGCCGGACGCGACCGCACGACGAAGATAGAGGT 1140  
Db 1081 GCGGCTGAGGTGATCTTACAAGAAAGCCGGACGCGACCGCACGACGAAGATAGAGGT 1140  
QY 1141 TACCGCATTCAGCGCGGACCGCATCACGCTCCGGCGGTGAGCTTGATTCGGACCGATC 1200  
Db 1141 TACCGCATTCAGCGCGGACCGCATCACGCTCCGGCGGTGAGCTTGATTCGGACCGATC 1200  
QY 1201 ATCGAGCCCGCTGAGTGTATGAGCTTTCAGGCGCTGTTGGACGGACGGGCGCGCAAG 1260  
Db 1201 ATCGAGCCCGCTGAGTGTATGAGCTTTCAGGCGCTGTTGGACGGACGGGCGCGCAAG 1260  
QY 1261 GGGCTTCCCGGGGCAAGCCATTTCTGTCGGCATGGACAAAGCTGTACTGCGAGTGGC 1320  
Db 1261 GGGCTTCCCGGGGCAAGCCATTTCTGTCGGCATGGACAAAGCTGTACTGCGAGTGGC 1320  
QY 1321 GCGGTCATGACTTCGAAGCGCGGGGGAAGAATCATCAAGACTCTTACCGCTGCCGTCGC 1380  
Db 1321 GCGGTCATGACTTCGAAGCGCGGGGGAAGAATCATCAAGACTCTTACCGCTGCCGTCGC 1380  
QY 1381 CGGAAGGTGTGTCGACCGCTCCGACCTGGGCGACGACGAAGGACAGCTGACGATG 1440  
Db 1381 CGGAAGGTGTGTCGACCGCTCCGACCTGGGCGACGACGAAGGACAGCTGACGATG 1440  
QY 1441 GCGGCACTGACACAAGTTCTGTCGGGAACGATCTTCAACAAGATTCAGGACGCGGAAGGC 1500  
Db 1441 GCGGCACTGACACAAGTTCTGTCGGGAACGATCTTCAACAAGATTCAGGACGCGGAAGGC 1500  
QY 1501 GACGAAGAGACGTTGGCGCTTCTGTGGGAAGCCCGCGACGCTTCGGCAAGCTCACTGAG 1560  
Db 1501 GACGAAGAGACGTTGGCGCTTCTGTGGGAAGCCCGCGACGCTTCGGCAAGCTCACTGAG 1560  
QY 1561 GCGCCTGAGAAGAGCGGCGAAGCGGCGAACCTTGTGGGAGCGCGCGACGCGCTGAAC 1620  
Db 1561 GCGCCTGAGAAGAGCGGCGAAGCGGCGAACCTTGTGGGAGCGCGCGACGCGCTGAAC 1620  
QY 1621 GCCCTTGAAGAGCTGTACGAAGACCGGCGGAGCGGCTGACGACGACCGCTGGCAGG 1680  
Db 1621 GCCCTTGAAGAGCTGTACGAAGACCGGCGGAGCGGCTGACGACGACCGCTGGCAGG 1680  
QY 1681 AAGCACTTCGGAAGCAACAGGACGCGCTGACGCTCCGGCAGCAAGGGCGGAAAGCGG 1740  
Db 1681 AAGCACTTCGGAAGCAACAGGACGCGCTGACGCTCCGGCAGCAAGGGCGGAAAGCGG 1740  
QY 1741 CTTGCCGAATTTGAAGCGCGGCAAGCCCGGAAGCTTCCCTTGACCAATGTTCCCGGAA 1800  
Db 1741 CTTGCCGAATTTGAAGCGCGGCAAGCCCGGAAGCTTCCCTTGACCAATGTTCCCGGAA 1800

[illegible]

Feature	Location/Qualifiers
RESULT 2	
ID	ABK12557 standard; DNA; 9880 BP.
XX	
AC	ABK12557;
XX	
DT	18-JUN-2002 (first entry)
XX	
DE	Plasmid pBPS EM051 T-DNA region.
XX	
KW	Phl C31 Integrase; recombinase; transgenic; plant;
KW	agricultural food product; self-excising polynucleotide;
KW	plasmid; ds; pBPS EM051; T-DNA; cyclic; circular.
XX	
OS	Synthetic.
XX	
FH	Key
FT	repeat_unit
FT	3..217
FT	/*tag= a
FT	/function= "The left T-DNA border and the right T-DNA border represent the end of the self-excising polynucleotide"
FT	
FT	misc_feature
FT	225..259
FT	/*tag= b
FT	/label= attB
FT	/note= "Recombinase target site"
FT	485..273
FT	/*tag= c
FT	/label= g7pa
FT	2288..519
FT	/*tag= d
FT	/product= "codA-tacCI translational fusion protein"
FT	2898..2303
FT	/*tag= e
FT	/label= Nopaline-synthase_promoter
FT	2925..3236
FT	/*tag= f
FT	/label= Octopine-synthase_promoter
FT	3260..4267
FT	/*tag= g
FT	/product= "tTA"
FT	4292..4558
FT	/*tag= h
FT	/label= Nopaline-synthase_terminator
FT	4597..4933
FT	/*tag= i
FT	/label= Top_10_promoter
FT	4977..7007
FT	/*tag= j
FT	/product= "Phl_C31_intINT"
FT	/note= "Novel phl C31 integrase"
FT	7027..7221
FT	/*tag= k
FT	/label= Octopine-synthase_terminator
FT	7253..8392
FT	/*tag= l
FT	
FT	promoter

FT		/label= Super_Promoter
FT	CDS	8413..9405
FT		/tag= m
FT		/product= "eGFP7INT"
FT	terminator	9411..9677
FT		/tag= n
FT		/label= Nopaline_synthase_terminator
FT	misc_feature	9690..9728
FT		/tag= O
FT		/label= attP
FT		/note= "Recombinase target site"
FT	repeat_unit	9735..9880
FT		/tag= p
FT		/function= "The right T-DNA border and the left T-DNA border represent the end of the self-excising polynucleotide"
FT		

WO200216609-A2.

28-FEB-2002.

27-AUG-2001; 2001WO-US26738.

25-AUG-2000; 2000US-227961P.

(BADI ) BASF PLANT SCI GMBH.  
(MCKE/) MCKERSIE B.

Mankin L;

WPI; 2002-280939/32.

New self-excising polynucleotides, useful for producing transgenic plants, removing transgenes from these plants or crops (e.g. food commodities), and restricting the distribution of transgenes within the environment

Example 2; Fig 8; 60pp; English.

The invention describes an isolated excisable polynucleotide, which comprises a desired tral polynucleotide and a recombinase polynucleotide operably linked to a promoter, all flanked by a pair of directly oriented recombination sites where the recombinase activity is regulatable. The self-excising polynucleotide is useful for producing transgenic plants, particularly for removing all transgenic nucleic acid sequences that encode a gene product from the transgenic plant, thus restoring the original genetic configuration of the genome. The polynucleotide is also useful in methods for eliminating unwanted nucleic acids in agricultural food products and for preventing the escape of certain transgenic traits into the environment. This sequence represents the plasmid pBPS EW051 T-DNA region used to test the self-excising DNA described in the invention.

Sequence 9880 BP; 2489 A; 2368 C; 2475 G; 2548 T; 0 other;

Query Match 100.0%; Score 2031; DB 24; Length 9880;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2031; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGGCACAAGGGGTGTGACCGGGGTGGATACGTAGTTCTGCTTCTACCTTTGATATA 60  
|||||  
4977 ATGGCACAAGGGGTGTGACCGGGGTGGATACGTAGTTCTGCTTCTACCTTTGATATA 5036  
|||||

61 TATATAATATATATCATTAATTAGTAGTAATATAATATTTCAAAATATTTTCAAAATA 120  
|||||  
5037 TATATAATATATATCATTAATTAGTAGTAATATAATATTTTCAAAATATTTTCAAAATA 5096  
|||||

121 AAGAATGTAGTATATAGCAATGCTTTCTGTAGTTTATTAAGTGTATATTTAAATT 180  
|||||  
5097 AAGAATGTAGTATATAGCAATGCTTTCTGTAGTTTATTAAGTGTATATTTAAATT 5156  
|||||

181 ATAACCTTTTCTAATATATATGACAAAAATTTGTTGATGTGACAGGTACGCGGGTCTTACGAC 240  
|||||

Db 5157 ATACTTTTCTAATATATATGACCAAAATTGTGTGATGTGAGGTACGGGGTCTTACGAC 5216  
QY 241 CGTACGTGCGGCGAGCGCGAGAAATTCGAGCGCAGCAAGCCCCAGCAGACACGCTAGCGCC 300  
Db 5217 CGTACGTGCGGCGAGCGCGAGAAATTCGAGCGCAGCAAGCCCCAGCAGACACGCTAGCGCC 5276  
QY 301 AACGAAGACAAGGCGCGGACCTTCAGCGCGAAGTCGAGCGCGAGGCGCGGCTTCAGG 360  
Db 5277 AACGAAGACAAGGCGCGGACCTTCAGCGCGAAGTCGAGCGCGAGCGCGGCGGCTTCAGG 5336  
QY 361 TTCGTGCGGCAATTCAGCGCAAGCGCGCGACGTCGCGCTTCGGGACGCGCGGAGCGCCCG 420  
Db 5337 TTCGTGCGGCAATTCAGCGCAAGCGCGCGACGTCGCGCTTCGGGACGCGCGGAGCGCCCG 5396  
QY 421 GAGTTCGAACGCATCTCTGAACGAATGCGCGCGCGCGGCTCAACATGATTCATTTCTAT 480  
Db 5397 GAGTTCGAACGCATCTCTGAACGAATGCGCGCGCGGCTCAACATGATTCATTTCTAT 5456  
QY 481 GACGTGTCGCGCTTCTCGCGCTGAAGGTCATGACGCGCATTCGATTTCTCGAATTG 540  
Db 5457 GACGTGTCGCGCTTCTCGCGCTGAAGGTCATGACGCGCATTCGATTTCTCGAATTG 5516  
QY 541 CTCGCCCTGGCGGTGACGATTTGTTTCCACTCAGAAAGCGCTCTCCGCGAGGAAACGTC 600  
Db 5517 CTCGCCCTGGCGGTGACGATTTGTTTCCACTCAGAAAGCGCTCTCCGCGAGGAAACGTC 5576  
QY 601 ATGACCTGATTCACCTGATTTATCGCGCTCAGCGCGCTCGCACAAGAATCTTCGCTGAAG 660  
Db 5577 ATGACCTGATTCACCTGATTTATCGCGCTCAGCGCGCTCGCACAAGAATCTTCGCTGAAG 5636  
QY 661 TCGCGGAAGATTCCTGACACGAGAAGACCTTCAGCGCGCAATTTGGCGGGTACGTGCGCGG 720  
Db 5637 TCGCGGAAGATTCCTGACACGAGAAGACCTTCAGCGCGCAATTTGGCGGGTACGTGCGCGG 5696  
QY 721 AAGCGGCTTACGCTTCGAGCTTGTTCGAGAGACGAAGAGATCACGCGCAACGCGCGA 780  
Db 5697 AAGCGGCTTACGCTTCGAGCTTGTTCGAGAGACGAAGAGATCACGCGCAACGCGCGA 5756  
QY 781 ATGTCATATGTCATCAACAAGCTTGGCAGCTCGACCACTCCCTTACCGGACCTTC 840  
Db 5757 ATGTCATATGTCATCAACAAGCTTGGCAGCTCGACCACTCCCTTACCGGACCTTC 5816  
QY 841 GAGTTCGAGCCCGACGTAATCCGGTGGTGGTGGCTGAGATCAAGACGACAAACACCTT 900  
Db 5817 GAGTTCGAGCCCGACGTAATCCGGTGGTGGTGGCTGAGATCAAGACGACAAACACCTT 5876  
QY 901 CCCTTCAAGCCCGGAGTCAAGCCCGCATTCACCCGGGAGCATCACGGGCTTGTAAAG 960  
Db 5877 CCCTTCAAGCCCGGAGTCAAGCCCGCATTCACCCGGGAGCATCACGGGCTTGTAAAG 5936  
QY 961 CGCATGGACGCTGACGCGCTGCCGACCGCGGAGAGACGATTTGGGAAGAACCGCTTCA 1020  
Db 5937 CGCATGGACGCTGACGCGCTGCCGACCGCGGAGAGACGATTTGGGAAGAACCGCTTCA 5996  
QY 1021 AGCGCTGGGACCGCGCAACCGTTATGCGAATCCTTCGGGACCGCGTATTTGGGGCTTC 1080  
Db 5997 AGCGCTGGGACCGCGCAACCGTTATGCGAATCCTTCGGGACCGCGTATTTGGGGCTTC 6056  
QY 1081 GCCGCTGAGGTGATCTACAAGAAGACCGGAGCGGACCGCGGACGACGAGATTGAGGGT 1140  
Db 6057 GCCGCTGAGGTGATCTACAAGAAGACCGGAGCGGACCGCGGACGACGAGATTGAGGGT 6116  
QY 1141 TACCGCATTCAGCGGACCGCATCACGCTCCGGCGGTGAGCTTGATTTGGGACCGATC 1200  
Db 6117 TACCGCATTCAGCGGACCGCATCACGCTCCGGCGGTGAGCTTGATTTGGGACCGATC 6176  
QY 1201 ATCGAGCCCGCTGAGTGTATGAGCTTCAGGCGTGTGGAGCGGAGGCGCGGCAAG 1260  
Db 6177 ATCGAGCCCGCTGAGTGTATGAGCTTCAGGCGTGTGGAGCGGAGGCGCGGCAAG 6236  
QY 1261 GGGCTTTCCCGGGGCAAGCATTTCTGTCGCGCATGGAACAAGCTGTACTGCGAGTGTGGC 1320  
Db 6237 GGGCTTTCCCGGGGCAAGCATTTCTGTCGCGCATGGAACAAGCTGTACTGCGAGTGTGGC 6296

QY 1321 GCCGTACGACTTCGAGCGCGGGGAGAAATCGATCAAGGACTCTTACCGCTGCCGTGC 1380  
Db 6297 GCCGTACGACTTCGAGCGCGGGGAGAAATCGATCAAGGACTCTTACCGCTGCCGTGC 6356  
QY 1381 CGGAAGGTGTCGACCGCTCCGACCTGGCGAGCACGAAGGACGTGCAACGTACAGATG 1440  
Db 6357 CGGAAGGTGTCGACCGCTCCGACCTGGCGAGCACGAAGGACGTGCAACGTACAGATG 6416  
QY 1441 GCGGCACTGACAACTTGTGCGGAAGCATCTTCAACAAGATCAGGACGCGCAAGGC 1500  
Db 6417 GCGGCACTGACAACTTGTGCGGAAGCATCTTCAACAAGATCAGGACGCGCAAGGC 6476  
QY 1501 GACGAAGACGTTGGCGCTTCTGTGGGAAGCCGCCGACGCTTCGGCAAGCTCAGTAG 1560  
Db 6477 GACGAAGACGTTGGCGCTTCTGTGGGAAGCCGCCGACGCTTCGGCAAGCTCAGTAG 6536  
QY 1561 GCGCCTGAGAAGACCGGGAACGCGGCAACCTTGTGCGGAGCGCGGAGCGCCCTGAAC 1620  
Db 6537 GCGCCTGAGAAGACCGGGAACGCGGCAACCTTGTGCGGAGCGCGGAGCGCCCTGAAC 6596  
QY 1621 GCCCTGAAGAGCTGTACGAAGACCGCGCGGAGGCGCGCTACGACGACCCGTTGGCAGG 1680  
Db 6597 GCCCTGAAGAGCTGTACGAAGACCGCGCGGAGGCGCGCTACGACGACCCGTTGGCAGG 6656  
QY 1681 AAGCACTTCGCGGAAGCAACAGGCAAGCGCTGACGCTCCGCGAGCAAGGGGCGGAAGCGG 1740  
Db 6657 AAGCACTTCGCGGAAGCAACAGGCAAGCGCTGACGCTCCGCGAGCAAGGGGCGGAAGCGG 6716  
QY 1741 CTGCGCACTGAAGCGCGCGGAGCGCGGAGGCTTCCCTTGACCAATGTTCCCGAA 1800  
Db 6717 CTGCGCACTGAAGCGCGCGGAGCGCGGAGGCTTCCCTTGACCAATGTTCCCGAA 6776  
QY 1801 GACGCGGAGCTGACCCGAGACCGGCTTAAGTGTGTGGGGCGCGGCTGAGTAGACGAC 1860  
Db 6777 GACGCGGAGCTGACCCGAGACCGGCTTAAGTGTGTGGGGCGCGGCTGAGTAGACGAC 6836  
QY 1861 AAGCGCTGTGCTCGGGCTTCTCTAGACAAGAATCGTTGTACGAAGTCGACTACGGGC 1920  
Db 6837 AAGCGCTGTGCTCGGGCTTCTCTAGACAAGAATCGTTGTGTACGAAGTCGACTACGGGC 6896  
QY 1921 AGGGGAGGGAAGCGCCCATCGAAGAGCGCGCTTCGATCACGTGGGCGAAGCCCGGACC 1980  
Db 6897 AGGGGAGGGAAGCGCCCATCGAAGAGCGCGCTTCGATCACGTGGGCGAAGCCCGGACC 6956  
QY 1981 GACGAGCAGGAAGACGAGCGCCAGGAGCGGACGCAAGAGCTAGCGGCGTAG 2031  
Db 6957 GACGAGCAGGAAGACGAGCGCCAGGAGCGGACGCAAGAGCTAGCGGCGTAG 7007

RESULT 3  
ABK12556  
ID ABK12556 standard; DNA; 2031 BP.  
XX  
AC ABK12556;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE DNA encoding novel phl C31 integrase, phlC31int\*INT.  
XX  
KW phl C31 integrase; recombinase; transgenic; plant;  
KW agricultural food product; self-excising polynucleotide;  
KW gene; ds; phlC31int\*INT.  
OS Synthetic.  
XX  
PN WO200216609-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 27-AUG-2001; 2001WO-US26738.  
XX  
PR 25-AUG-2000; 2000US-227961P.

XX (BADI ) BASF PLANT SCI GMBH.  
PA (MCKE/) MCKERSIE B.  
XX  
XX Mankin L;  
XX WPI; 2002-280939/32.  
DR  
XX  
XX New self-excising polynucleotides, useful for producing transgenic  
PT plants, removing transgenes from these plants or crops (e.g. food  
PT commodities), and restricting the distribution of transgenes within the  
PT environment  
XX  
XX Claim 5; Fig 7; 60pp; English.  
XX  
XX The invention describes an isolated excisable polynucleotide, which  
CC comprises a desired trait polynucleotide and a recombinase polynucleotide  
CC operably linked to a promoter, all flanked by a pair of directly oriented  
CC recombination sites where the recombinase activity is regulatable. The  
CC self-excising polynucleotide is useful for producing transgenic plants,  
CC particularly for removing all transgenic nucleic acid sequences that  
CC encode a gene product from the transgenic plant, thus restoring the  
CC original genetic configuration of the genome. The polynucleotide is also  
CC useful in methods for eliminating unwanted nucleic acids in agricultural  
CC food products and for preventing the escape of certain transgenic traits  
CC into the environment. This sequence encodes the novel phic31 integrase  
CC phic31int\*INT, created for use in the self-excising polynucleotide  
CC described in the invention.  
XX  
SQ Sequence 2031 BP; 454 A; 558 C; 640 G; 379 T; 0 other;  
  
Query Match 99.2%; Score 2015; DB 24; Length 2031;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 2021; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

DB 541 CTCGCCCTGGCGGTGACGATTTGTTCCACTCAGGAAGCGCTTCCGGCAGGAAACGTC 600  
QY 601 ATGACCTGATTCACCTGATTATGCGGCTGCAGCGCTGCACAAAGAAATCTTGGCTGAG 660  
DB 601 ATGACCTGATTCACCTGATTATGCGGCTGCAGCGCTGCACAAAGAAATCTTGGCTGAG 660  
QY 661 TCGCGAAGATTCGACACGAGAAGACCTTCAGCGGAATTGGCGGGGTACGTCGGGG 720  
DB 661 TCGCGAAGATTCGACACGAGAAGACCTTCAGCGGAATTGGCGGGGTACGTCGGGG 720  
QY 721 AAGCGCCTTACGGCTTCGAGCTTGTTTCGAGACGAGAAGACATCACCGCCACGCCGA 780  
DB 721 AAGCGCCTTACGGCTTCGAGCTTGTTTCGAGACGAGAAGACATCACCGCCACGCCGA 780  
QY 781 ATGCTCAATGTGTCATCAACAAGCTTGGCCACTCGACCACCTCCCTTACCGGACCTTC 840  
DB 781 ATGCTCAATGTGTCATCAACAAGCTTGGCCACTCGACCACCTCCCTTACCGGACCTTC 840  
QY 841 GAGTTCAGCCGACGTAATCCGCTGCTGCGGTGAGATCAAGCAGCAACACCTT 900  
DB 841 GAGTTCAGCCGACGTAATCCGCTGCTGCGGTGAGATCAAGCAGCAACACCTT 900  
QY 901 CCCTTCAAGCCGGGACGTCAAGCCGCCATTACCCGGGACAGATCAAGGGCTTTGTAAG 960  
DB 901 CCCTTCAAGCCGGGACGTCAAGCCGCCATTACCCGGGACAGATCAAGGGCTTTGTAAG 960  
QY 961 CGCATGACCGCTGACCGCCGTGCCGACCCGGGGCGAGACGATTGGGAAGAAGACCGCTTCA 1020  
DB 961 CGCATGACCGCTGACCGCCGTGCCGACCCGGGGCGAGACGATTGGGAAGAAGACCGCTTCA 1020  
QY 1021 AGCGCCTGGGACCCGGGCAACCGTTATGCGAATCCTTCGGGACCCGCTATTCGGGCTTC 1080  
DB 1021 AGCGCCTGGGACCCGGGCAACCGTTATGCGAATCCTTCGGGACCCGCTATTCGGGCTTC 1080  
QY 1081 GCCGCTGAGTGATCTACAAGAAGAGCGGACGCGACGCGCCGACCAAGAAATTGAGGT 1140  
DB 1081 GCCGCTGAGTGATCTACAAGAAGAGCGGACGCGACGCGCCGACCAAGAAATTGAGGT 1140  
QY 1141 TACCGCATTCAGCGCGGACCCGATCACGCTCCGCGGTGAGCTTGAATGCGACCGATC 1200  
DB 1141 TACCGCATTCAGCGCGGACCCGATCACGCTCCGCGGTGAGCTTGAATGCGACCGATC 1200  
QY 1201 ATCGAGCCCGCTGAGTGTATGAGCTTCAGCGCTGTTGAGCGGACAGGGGCGCGCAAG 1260  
DB 1201 ATCGAGCCCGCTGAGTGTATGAGCTTCAGCGCTGTTGAGCGGACAGGGGCGCGCAAG 1260  
QY 1261 GGGCTTCCCGGGGCAAGCCATTCTGTCCGCCATGGACAAGCTGTACTGCGAGTGTG 1320  
DB 1261 GGGCTTCCCGGGGCAAGCCATTCTGTCCGCCATGGACAAGCTGTACTGCGAGTGTG 1320  
QY 1321 GCCGTGATGACTTCGAAGCGGGGGAAGATCGATCAAGGACTCTTACCGCTGCCGTGC 1380  
DB 1321 GCCGTGATGACTTCGAAGCGGGGGAAGATCGATCAAGGACTCTTACCGCTGCCGTGC 1380  
QY 1381 CGGAAGTGTGTCAGCCGTCGACCTGGGACAGCAGCAAGGACGTGCAACGTCAGCATG 1440  
DB 1381 CGGAAGTGTGTCAGCCGTCGACCTGGGACAGCAGCAAGGACGTGCAACGTCAGCATG 1440  
QY 1441 GCGGCACTGACAAGTTGCTTGGGAAAGCATCTTCAACAAGATCAAGCAGCGCAAGGC 1500  
DB 1441 GCGGCACTGACAAGTTGCTTGGGAAAGCATCTTCAACAAGATCAAGCAGCGCAAGGC 1500  
QY 1501 GACGAAGAGAGCTTGGCGCTTCTGTGGGAAGCCGCGCGACGCTTCCGCAAGCTCACTAG 1560  
DB 1501 GACGAAGAGAGCTTGGCGCTTCTGTGGGAAGCCGCGCGACGCTTCCGCAAGCTCACTAG 1560  
QY 1561 GCGCCTGAGAAGAGCGGGCAAGCGGGCAACCTTGTGCGGAGCGCGCGAGCCCTGAAC 1620  
DB 1561 GCGCCTGAGAAGAGCGGGCAAGCGGGCAACCTTGTGCGGAGCGCGCGAGCCCTGAAC 1620  
QY 1621 GCCCTGAAGAGCTGTACGAAGACCGCGGACAGGCGCGTACGACGAGCCGTTGGCAGG 1680  
DB 1621 GCCCTGAAGAGCTGTACGAAGACCGCGGACAGGCGCGTACGACGAGCCGTTGGCAGG 1680



QY 1681 AAGCACTTCCGGAAGCAACAGGACGCTGACGCTCCGGCAGCAAGGGCCGGAAGAGCGG 1740  
DB 1681 AAGCACTTCCGGAAGCAACAGGACGCTGACGCTCCGGCAGCAAGGGCCGGAAGAGCGG 1740  
QY 1741 CTTGCCGAAGCTTGAAGCCGCGCAAGCCCGAAGCTTCCCTTGACCAATGTTCCCGAA 1800  
DB 1741 CTTGCCGAAGCTTGAAGCCGCGCAAGCCCGAAGCTTCCCTTGACCAATGTTCCCGAA 1800  
QY 1801 GAGCGCGACGCTGACCCGACGCGCCCTAAGTCGTGTGGGGGGCGCGCTCAGTAGACGAC 1860  
DB 1801 GAGCGCGACGCTGACCCGACGCGCCCTAAGTCGTGTGGGGGGCGCGCTCAGTAGACGAC 1860  
QY 1861 AAGCGCGTGTCTGTCGGGCTTCTGACAGACATGCTGTACAGGAAGTCGACTACGGGC 1920  
DB 1861 AAGCGCGTGTCTGTCGGGCTTCTGACAGACATGCTGTACAGGAAGTCGACTACGGGC 1920  
QY 1921 AGGGGCGAGGGAAGCGCCCATCAGAGAGCGCGCTTCATCAGTGGCGGAAGCCCGGAC 1980  
DB 1921 AGGGGCGAGGGAAGCGCCCATCAGAGAGCGCGCTTCATCAGTGGCGGAAGCCCGGAC 1980  
QY 1981 GACGACGACGAGAGACGACGCGCCAGAGCGGACGAGAGAGAGAGAGAGAGAGAGAGAG 2031  
DB 1981 GACGACGACGAGAGAGACGACGCGCCAGAGCGGACGAGAGAGAGAGAGAGAGAGAGAGAG 2031

RESULT 4  
AAD04935

ID AAD04935 standard; DNA; 5711 BP.

AC AAD04935;

DT 17-JUL-2001 (first entry)

XX C31-Int expression vector pRK65 for phic31 integrase.

XX Gene trapping construct; conditional mutation; unidirectional inversion;  
KW recombinase recognition sequence; RRS; disruption cassette;  
KW selection cassette; transgenic organism; expression vector pRK65;  
XX phic31 integrase; C31-Int mediated inversion; ds.

OS Chimeric - Bacteriophage phi-C31.  
OS Chimeric - Cytomegalovirus.  
XX Chimeric - Unidentified.

Key Location/Qualifiers

FT promoter 1..700

FT /tag= a

FT /note= "Cytomegalovirus immediate early gene promoter"

FT Intron 700..970

FT /tag= b

FT /note= "Hybrid Intron"

FT CDS 978..2819

FT /tag= c

FT /product= "C31-Int, a phic31 phage derived integrase"

FT misc\_feature 2831..3020

FT /tag= d

FT /note= "Polyadenylation sequence"

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

WO200129208-A1.

26-APR-2001.

16-OCT-2000; 2000WO-EP10162.

16-OCT-1999; 99EP-0120592.

27-OCT-1999; 99US-0162016.

(ARTE-) ARTEMIS PHARM GMBH.  
(FRAN-) FRANKGEN BIOTECHNOLOGIE AG.

Kuehn R, Von Melchener H, Altschmied J;

DR WPI; 2001-308486/32.

XX New gene trapping construct capable of causing conditional mutations in  
PT genes, comprises functional DNA segment inserted in sense or antisense  
PT direction relative to gene to be trapped  
PS Example 2; Page 61-63; 78pp; English.

CC The present invention relates to a conditional gene trapping construct  
CC capable of causing conditional mutations in genes. The gene trapping  
CC construct comprises two functional DNA segments, each being flanked by  
CC two recombinase recognition sequences (RRS) specific to site specific  
CC recombinase which is capable of unidirectional inversion of double  
CC standard DNA segment. One of the DNA segment (disruption cassette) is  
CC inserted in antisense orientation relative to the transcriptional  
CC orientation of the gene to be trapped. The other DNA segment (selection  
CC cassette) is inserted in sense direction relative to the transcriptional  
CC orientation of the gene to be trapped. The cell comprising the gene  
CC trapping construct is useful for the identification and/or isolation of  
CC genes. The transgenic organism comprising the gene trapping construct is  
CC useful to study gene function at various developmental stages. The gene  
CC trapping construct is useful for mutationally inactivating all cellular  
CC genes. The present sequence is an expression vector pRK65. The pRK65  
CC vector is transfected into a murine cell along with the test vector  
CC pRK73. The pRK65 is used for the expression of phic31 integrase to  
CC detect C31-Int (a phic31 phage derived integrase) mediated inversion.

CC Sequence 5711 BP; 1329 A; 1512 C; 1616 G; 1254 T; 0 other;

Query Match 89.2%; Score 1812.6; DB 22; Length 5711;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1815; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 213 GATGTGACGTAAGCGGGGTGCTTACGACCGTCACTGCGCGGAGCGGAGAAATTCAGCGC 272  
DB 1001 GGTGACAGCTACCGGGGTGCTTACGACCGTCACTGCGCGGAGCGGAGAAATTCAGCGC 1060  
QY 273 AGCAAGCCGACGACACAGCGGTAGCGCCCAAGCAAGAGCGGCGGACCTTCAGCGCGA 332  
DB 1061 AGCAAGCCGACGACACAGCGGTAGCGCCCAAGCAAGAGCGGCGGACCTTCAGCGCGA 1120  
QY 333 AGTCAGCGCGAGCGGGGCGGCTTACAGTTCGTCGGGCAATTCAGCGAGCGCGGCGAC 392  
DB 1121 AGTCAGCGCGAGCGGGGCGGCTTACAGTTCGTCGGGCAATTCAGCGAGCGCGGCGAC 1180  
QY 393 GTCGGGCTTCGGGAGCGCGGCGGCGGAGTTCGCAACGCAATTCGCAAGATGCGGCGC 452  
DB 1181 GTCGGGCTTCGGGAGCGCGGCGGCGGAGTTCGCAACGCAATTCGCAAGATGCGGCGC 1240  
QY 453 CGGGCGCTCAACATGATATGTTATGACGCTGCGGCTTCGCGGCTGAAGGTCAAT 512  
DB 1241 CGGGCGCTCAACATGATATGTTATGACGCTGCGGCTTCGCGGCTGAAGGTCAAT 1300  
QY 513 GGACGCGATTCGATGCTCTCGGAATTCGCGGCTGGCGGTGACGATGTTCCACTCA 572  
DB 1301 GGACGCGATTCGATGCTCTCGGAATTCGCGGCTGGCGGTGACGATGTTCCACTCA 1360  
QY 573 GGAAGGCGTCTTCGGGAGGGAACGTCATGCACTGATTCACCTGATTAATGCGGCTCA 632  
DB 1361 GGAAGGCGTCTTCGGGAGGGAACGTCATGCACTGATTCACCTGATTAATGCGGCTCA 1420  
QY 633 CGCGTCGACAAAGATCTTCGCTGAAGTCGGCGGAAGATTCGACACGAGAAGACTTCA 692  
DB 1421 CGCGTCGACAAAGATCTTCGCTGAAGTCGGCGGAAGATTCGACACGAGAAGACTTCA 1480  
QY 693 GCGGCAATTCGGGCGGTACGTCGGCGGGAAGCGGCTTACGGCTTCGAGCTTGTTCGGA 752  
DB 1481 GCGGCAATTCGGGCGGTACGTCGGCGGGAAGCGGCTTACGGCTTCGAGCTTGTTCGGA 1540  
QY 753 GACGAAGAGATACGCGCAAGCGGCGCAATGCTCAATGCTCATCAACAAGCTTGGCGA 812  
DB 1541 GACGAAGAGATACGCGCAAGCGGCGCAATGCTCAATGCTCATCAACAAGCTTGGCGA 1600



QY 813 CTGACCACTCCCTTACCAGCCCTTTCGAGTTCGAGCCCGACGTAATCCGGTGGTG 872  
Db 1601 CTCGACCACTCCCTTACCAGCCCTTTCGAGTTCGAGCCCGACGTAATCCGGTGGTG 1660  
QY 873 GCGTAGATCAAGACGACACAACACCTTCCCTTCAAGCCGGGAGTCAAGCCGCATTCA 932  
Db 1661 GCGTAGATCAAGACGACACAACACCTTCCCTTCAAGCCGGGAGTCAAGCCGCATTCA 1720  
QY 933 CCGGGCAGCATCAGGGGGCTTTGTAAGCGCATGAGCGCTGACGCCGTGCCGACCCGGGG 992  
Db 1721 CCGGGCAGCATCAGGGGGCTTTGTAAGCGCATGAGCGCTGACGCCGTGCCGACCCGGGG 1780  
QY 993 CGAGACGATTGGGAAGAAGACCCCTTCAAGCGCCCTGGGACCCGGCAACCGTTATGCGAAT 1052  
Db 1781 CGAGACGATTGGGAAGAAGACCCCTTCAAGCGCCCTGGGACCCGGCAACCGTTATGCGAAT 1840  
QY 1053 CCTTCGGGACCCCGCGTATTGGGGCTTCGCCGCTGAGTGATCTACAAGAAGACCCGGA 1112  
Db 1841 CCTTCGGGACCCCGCGTATTGGGGCTTCGCCGCTGAGTGATCTACAAGAAGACCCGGA 1900  
QY 1113 CGGACGCCGACCAAGAGATTGAGGTTACCAGCATTCAGCCGACCCGATGACGCTCCG 1172  
Db 1901 CGGACGCCGACCAAGAGATTGAGGTTACCAGCATTCAGCCGACCCGATGACGCTCCG 1960  
QY 1173 GCCGTCGAGCTTGATTGCGGACCGATCATCGAGCCGCTGAGTGTATGAGCTTCAGGC 1232  
Db 1961 GCCGTCGAGCTTGATTGCGGACCGATCATCGAGCCGCTGAGTGTATGAGCTTCAGGC 2020  
QY 1233 GTGTTGGACGGGACAGGGGGCGGCAAGGGGCTTCCCGGGGGCAAGCCATTCTGCCG 1292  
Db 2021 GTGTTGGACGGGACAGGGGGCGGCAAGGGGCTTCCCGGGGGCAAGCCATTCTGCCG 2080  
QY 1293 CATGACAAGCTGTACTGCGAGTGTGGCGCCGTCATGACTTCGAAGCGCGGGAAGAATC 1352  
Db 2081 CATGACAAGCTGTACTGCGAGTGTGGCGCCGTCATGACTTCGAAGCGCGGGAAGAATC 2140  
QY 1353 GATCAAGGACTCTTACCGCTGCGCTGCCGGAAGTGTGACCCGTCGACCTGAGCA 1412  
Db 2141 GATCAAGGACTCTTACCGCTGCGCTGCCGGAAGTGTGACCCGTCGACCTGAGCA 2200  
QY 1413 GCACGAGGACGTCGACAGCTGACATGGCGGACATGACAAATTCTGCGGAAGCAT 1472  
Db 2201 GCACGAGGACGTCGACAGCTGACATGGCGGACATGACAAATTCTGCGGAAGCAT 2260  
QY 1473 CTTCAACAAGATCAGGCAAGCCGGAAGGCGAGCAAGACGTTGGCGTTCTGTGGAAGC 1532  
Db 2261 CTTCAACAAGATCAGGCAAGCCGGAAGGCGAGCAAGACGTTGGCGTTCTGTGGAAGC 2320  
QY 1533 CGCCGACGCTTGGCAAGCTCAGTGAAGCGCTGAGAAGAGCGGGAACGGGCGAACCT 1592  
Db 2321 CGCCGACGCTTGGCAAGCTCAGTGAAGCGCTGAGAAGAGCGGGAACGGGCGAACCT 2380  
QY 1593 TGTTCGGGAGCGCGCCGACGCCCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGGC 1652  
Db 2381 TGTTCGGGAGCGCGCCGACGCCCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGGC 2440  
QY 1653 AGGCGGTACGACGAGCAGCCGTTGGCAGGAAGCATTCCGGAAGCAACAGGACGCGTGAC 1712  
Db 2441 AGGCGGTACGACGAGCAGCCGTTGGCAGGAAGCATTCCGGAAGCAACAGGACGCGTGAC 2500  
QY 1713 GCTCCGCGACGAAGGGGCGAAGAGCGGCTTGCCGAACCTGAAGCCGCCGAGCCCGAA 1772  
Db 2501 GCTCCGCGACGAAGGGGCGAAGAGCGGCTTGCCGAACCTGAAGCCGCCGAGCCCGAA 2560  
QY 1773 GCTTCCTTGAACCAATGTTCCCGAAGACGCCGACGCTGACCCGACCGGCTTAAGTC 1832  
Db 2561 GCTTCCTTGAACCAATGTTCCCGAAGACGCCGACGCTGACCCGACCGGCTTAAGTC 2620  
QY 1833 GTGCTGGGGGGCGCGCTCAGTAGACGACAAGCGCTGTTCGCTGGGCTCTTCGTAGACAA 1892  
Db 2621 GTGCTGGGGGGCGCGCTCAGTAGACGACAAGCGCTGTTCGCTGGGCTCTTCGTAGACAA 2680  
QY 1893 GATGTTGTACGAAGTGCATACGGGCGAGGGGCGAAGCGCCCATCGAGAAGCGCGC 1952

Db 2681 GATCGTTGTACGAAGTGCATACGGGCGAGGGGCGAGGGAGGCCATCGAGAAGCGCGC 2740  
QY 1953 TTGATCACGTGGGCGAAGCCGCCGACGACGAGCAGGAAGACGAGCCGAGAGCGCAC 2012  
Db 2741 TTGATCACGTGGGCGAAGCCGCCGACGACGAGCAGGAAGACGAGCCGAGAGCGCAC 2800  
QY 2013 GGAAGACGTAGCGCGTAG 2031  
Db 2801 GGAAGACGTAGCGCGTAG 2819  
RESULT 5  
AAH74879  
ID AAH74879 standard; DNA; 1842 BP.  
XX  
AC AAH74879;  
XX  
DT 29-OCT-2001 (first entry)  
XX  
DE Nucleotide sequence of a wildtype bacteriophage thetaC31 integrase.  
XX  
KW Recombinase; genetic engineering; monogenic disorder; ADA deficiency;  
KW cystic fibrosis; familial-hypercholesterolaemia; haemophilia;  
KW chronic granulomatous disease; Duchenne's muscular dystrophy;  
KW Fanconi's disease; anemia; sickle-cell anemia; Gaucher's disease;  
KW Hunter's syndrome; X-linked severe combined immunodeficiency; SCID;  
KW infectious disease; acquired disorder; tumour; cancer; integrase; ss.  
XX  
OS Bacteriophage thetaC31.  
XX  
PN WO200161049-A1.  
XX  
PD 23-AUG-2001.  
XX  
PF 16-FEB-2001; 2001WO-US05269.  
XX  
PR 18-FEB-2000; 2000US-0183759.  
XX  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
XX  
PI Calos MP, Scilmentl CR;  
XX  
DR WPI; 2001-522610/57.  
XX  
PT Identifying altered recombinase, involves hybridizing recombination sites in cells having specific coding sequence, transformed with altered recombinase gene, and isolating cells having product of the sequence -  
XX  
PS Disclosure; Fig 4; 101pp; English.  
XX  
CC The specification describes a method for identifying altered recombinases. The method comprises transforming cells with a first plasmid comprising two recombination sites and a coding sequence of interest, and a second plasmid encoding an altered recombinase; allowing recombination of the two recombination sites using the altered recombinase; screening and isolating transformed cells comprising the product of the sequence of interest; and identifying the altered recombinase. The altered recombinase is useful for site-specifically integrating a polynucleotide sequence of interest in a genome of a cell. It is also useful in genetic engineering of chromosomes of higher cells, and for the generation of transgenic cells, tissues, plants and animals. The altered recombinase is useful for treating monogenic disorders, e.g. ADA deficiency, cystic fibrosis, familial-hypercholesterolaemia, anemia, chronic granulomatous disease, Duchenne's muscular dystrophy, haemophilia, Fanconi's disease, sickle-cell anemia, Gaucher's disease, Hunter's syndrome and X-linked severe combined immunodeficiency (SCID), infectious diseases including viral and bacterial infections, acquired disorders including solid tumours and haematopoietic cancers such as leukaemias and lymphomas, and other cancers. The present sequence encodes a wild type bacteriophage thetaC31 integrase.  
XX

SQ Sequence 1842 BP; 387 A; 547 C; 616 G; 292 T; 0 other;

Query Match 89.28; Score 1811; DB 22; Length 1842;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1814; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 213 GATGTGCAGGTACGGGGTCTTACGACCGTCAGTCGCGGAGCGGAGAATTGAGCGC 272  
Db 24 GTTGACACAGTACGGGGTCTTACGACCGTCAGTCGCGGAGCGGAAAATTGAGCGC 83  
QY 273 AGCAAGCCCGACGACAGCGGTAGCGCCAAAGAGACAAAGCGCGGACCTTCAGCGCGA 332  
Db 84 AGCAAGCCCGACGACAGCGGTAGCGCCAAAGAGACAAAGCGCGGACCTTCAGCGCGA 143  
QY 333 AGTCGAGCGCGAGCGGGCGCGGTTACAGTTCGTGCGGCAATTCAGCGGAGCGCGGCGAC 392  
Db 144 AGTCGAGCGCGAGCGGGCGCGGTTACAGTTCGTGCGGCAATTCAGCGGAGCGCGGCGAC 203  
QY 393 GTGCGCGTTCGAGCGCGGAGCGCGCGGAGTTCGACGCAATCCTGAACGAATGCCGCGC 452  
Db 204 GTGCGCGTTCGAGCGCGGAGCGCGCGGAGTTCGACGCAATCCTGAACGAATGCCGCGC 263  
QY 453 CGGGCGGCTCAACATGATCATTTCTGATGACGTGTGCGCGTTCGCGCTGAAGGTGAT 512  
Db 264 CGGGCGGCTCAACATGATCATTTCTGATGACGTGTGCGCGTTCGCGCTGAAGGTGAT 323  
QY 513 GGACCGGATCCGATTTCTCTCGGAATGCTCGCCCTGGCGGTGACGATTTGTTCCACTCA 572  
Db 324 GGACCGGATCCGATTTCTCTCGGAATGCTCGCCCTGGCGGTGACGATTTGTTCCACTCA 383  
QY 573 GGAAGCGCTTCCTCCGAGGAGGAAACGTCATGACCTGATTCACCTGATTAAGCGCTCGA 632  
Db 384 GGAAGCGCTTCCTCCGAGGAGGAAACGTCATGACCTGATTCACCTGATTAAGCGCTCGA 443  
QY 633 CGCGTCGACAAAGATCTTCGCTGAAGTCGCGGAAGATTCGACACGAAGAACCTTCA 692  
Db 444 CGCGTCGACAAAGATCTTCGCTGAAGTCGCGGAAGATTCGACACGAAGAACCTTCA 503  
QY 693 GCGGCAATTTGGGCGGGTACGTCGCGGGGAAAGCGCTTACGCGCTTTCGAGCTTTTCGGA 752  
Db 504 GCGGCAATTTGGGCGGGTACGTCGCGGGGAAAGCGCTTACGCGCTTTCGAGCTTTTCGGA 563  
QY 753 GACGAAGAGATCAGCGCAACGCGCAATGTCATGTCATCAACAAGCTTGCGCA 812  
Db 564 GACGAAGAGATCAGCGCAACGCGCAATGTCATGTCATCAACAAGCTTGCGCA 623  
QY 813 CTGACCACTCCCTTACCGGACCTTCGAGTTCGAGCCCGACGTAATCCGGTGTGTG 872  
Db 624 CTGACCACTCCCTTACCGGACCTTCGAGTTCGAGCCCGACGTAATCCGGTGTGTG 683  
QY 873 GCGTGAGATCAAGACGACAAACACCTTCCCTTCAAGCCGGGACGTCAGCCGCCATTC 932  
Db 684 GCGTGAGATCAAGACGACAAACACCTTCCCTTCAAGCCGGGACGTCAGCCGCCATTC 743  
QY 933 CCGGCGACGATCAGCGGGCTTTGTAAGCGCATGACGCGCTGACGCGCGGACCGGGG 992  
Db 744 CCGGCGACGATCAGCGGGCTTTGTAAGCGCATGACGCGCTGACGCGCGGACCGGGG 803  
QY 993 CGAGACGATTGGGAAGAAGACCGCTTCAAGCGCTGGGACCGGCAACCGTTATGCGAAT 1052  
Db 804 CGAGACGATTGGGAAGAAGACCGCTTCAAGCGCTGGGACCGGCAACCGTTATGCGAAT 863  
QY 1053 CCTTGGGAGACCGCGTATTCGGGGCTTCGCGCTGAGGTGATCTACAAGAAGACCGGGA 1112  
Db 864 CCTTGGGAGACCGCGTATTCGGGGCTTCGCGCTGAGGTGATCTACAAGAAGACCGGGA 923  
QY 1113 CGGACCGCCGACGACGAGATTGAGGTTACCGCATTCAGCGGACCGCGATCAGCTCCG 1172  
Db 924 CGGACCGCCGACGACGAGATTGAGGTTACCGCATTCAGCGGACCGCGATCAGCTCCG 983  
QY 1173 GCGGCTGAGCTTGAATGCGGACCGATCATGACCGCGTGAAGTGAAGCTTCAGGC 1232  
Db 984 GCGGCTGAGCTTGAATGCGGACCGATCATGACCGCGTGAAGTGAAGCTTCAGGC 1043

QY 1233 GTGTTGACGCGAGGGGGCGCGGCAAGGGCTTTCCCGGGGCAAGCCATTCTGTCCG 1292  
Db 1044 GTGTTGACGCGAGGGGGCGCGGCAAGGGCTTTCCCGGGGCAAGCCATTCTGTCCG 1103  
QY 1293 CATGACAAAGCTGTACTGCGAGTGTGGCGCGGTATGACTTCGAAGCGGGGAGAATC 1352  
Db 1104 CATGACAAAGCTGTACTGCGAGTGTGGCGCGGTATGACTTCGAAGCGGGGAGAATC 1163  
QY 1353 GATCAAGACTCTTACCGCTGCGCTGCGCGGAGAGGTGTGACCCCGTCCGACTGGGCA 1412  
Db 1164 GATCAAGACTCTTACCGCTGCGCTGCGCGGAGAGGTGTGACCCCGTCCGACTGGGCA 1223  
QY 1413 GCACGAAGCGACGTGCAACGTGACGATGGCGGACACTGACAAAGTTGTTGGCAACGAT 1472  
Db 1224 GCACGAAGCGACGTGCAACGTGACGATGGCGGACACTGACAAAGTTGTTGGCAACGAT 1283  
QY 1473 CTCAACAGATCAGGACCGCCGAGGCGGACGAGAGACGTTGGCGTTGTGGGAAGC 1532  
Db 1284 CTCAACAGATCAGGACCGCCGAGGCGGACGAGAGACGTTGGCGTTGTGGGAAGC 1343  
QY 1533 CGCCCGACGCTTCGCAAGCTCACTGAGGCGGCTGAGAAAGCGGCGAAGCGGCAACCT 1592  
Db 1344 CGCCCGACGCTTCGCAAGCTCACTGAGGCGGCTGAGAAAGCGGCGAAGCGGCAACCT 1403  
QY 1593 TGTTCGCGAGCGGCGCGGACCGCTGAACGCGCTTGAAGAGCTGTACGAAGACCGCGGC 1652  
Db 1404 TGTTCGCGAGCGGCGCGGACCGCTGAACGCGCTTGAAGAGCTGTACGAAGACCGCGGC 1463  
QY 1653 AGCGCGTACGACGCGACCGCTTGGCGAGGAGACACTTCGGAAGCAACAGCGCGCTGAC 1712  
Db 1464 AGCGCGTACGACGCGACCGCTTGGCGAGGAGACACTTCGGAAGCAACAGCGCGCTGAC 1523  
QY 1713 GCTCCGCGACGCAAGGGCGGAGAGCGGCTTGCCTGAAGAGCGGCGGAAAGCCCGGAA 1772  
Db 1524 GCTCCGCGACGCAAGGGCGGAGAGCGGCTTGCCTGAAGAGCGGCGGAAAGCCCGGAA 1583  
QY 1773 GCTTCCCTTGAACCAATGTTCCCGGAAGACGCGGACGCTGACCGGCGGCTTAAGTC 1832  
Db 1584 GCTTCCCTTGAACCAATGTTCCCGGAAGACGCGGACGCTGACCGGCGGCTTAAGTC 1643  
QY 1833 GTGTTGGGGCGCGCGTCACTAGACGACAGCGCGTGTCTGCGGCTCTTCTGAGACAA 1892  
Db 1644 GTGTTGGGGCGCGCGTCACTAGACGACAGCGCGTGTCTGCGGCTCTTCTGAGACAA 1703  
QY 1893 GATCTTGTCAAGATGCACTACGCGGACGCGGAGGAGCGGCGGCAAGAGCGGCGC 1952  
Db 1704 GATCTTGTCAAGATGCACTACGCGGACGCGGAGGAGCGGCGGCAAGAGCGGCGC 1763  
QY 1953 TTCGATCAGTGGGCGAAGCGCGGACCGGACGACGAGAGAGAGCGGCGGAGAGCGGAC 2012  
Db 1764 TTCGATCAGTGGGCGAAGCGCGGACCGGACGACGAGAGAGAGCGGCGGAGAGCGGAC 1823  
QY 2013 GGAAGAGTAGCGGCGTAG 2031  
Db 1824 GGAAGAGTAGCGGCGTAG 1842

RESULT 6

AA006904

AA006904 standard; DNA; 3401 BP.

XX AA006904;

XX 06-MAR-1991 (first entry)

XX Sequence encoding site specific integrating function of actinomycete  
DE phage phi C31.

XX Antibiotic production; isovaleryl; spiramycin; ds.

XX Page phi C31.







Db	924	CGGCACGCCGACCACGAACTTGAGGGTTACCCGATTCAGCGCCGACCCGATCACGCTCCG	983
QY	1173	GCCCGTCGAGCTTGATTGCCGACCCGATCATCGAGCCCGCTGAGTGGTATGAGCTTCAGGC	1232
Db	984	GCCGGTCGAGCTTGATTGCCGACCCGATCATCGAGCCCGCTGAGTGGTATGAGCTTCAGGC	1043
QY	1233	GTGGTTGACGGCAGGGGGCGCGCAAGGGGCTTTCGCGGGGGCAAGCCATTCTGTCCG	1292
Db	1044	GTGCTTGACGGCAGGGGGCGCGCAAGGGGCTTTCGCGGGGGCAAGCCATTCTGTCCG	1103
QY	1293	CATGGACAAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAAGCGCGGGGAAGATC	1352
Db	1104	CATGGACAAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAAGCGCGGGGAAGATC	1163
QY	1353	GATCAAGGACTCTTACCCTGCCGTCGCCGGAAAGGTGGTTCGACCCGCTCCGACCTGGCA	1412
Db	1164	GATCAAGGACTCTTACCCTGCCGTCGCCGGAAAGGTGGTTCGACCCGCTCCGACCTGGCA	1223
QY	1413	GCACGAAGGCACGTGCACGCTCAGCATGGCGGCACTCGACAAGTTCTGTGCGGAACGAT	1472
Db	1224	GCACGAAGGCACGTGCACGCTCAGCATGGCGGCACTCGACAAGTTCTGTGCGGAACGAT	1283
QY	1473	CTTCAACAAGATCAGGCACGCCGGAAGCGCAGCAAGAGAGCTTGGCGCTTCTGTGGGAAGC	1532
Db	1284	CTTCAACAAGATCAGGCACGCCGGAAGCGCAGCAAGAGAGCTTGGCGCTTCTGTGGGAAGC	1343
QY	1533	CGCCCGACGCTTCGGCAAGCTCACTGAGGGCGCTGAGAAGAGCGGGCAACGCGCAACT	1592
Db	1344	CGCCCGACGCTTCGGCAAGCTCACTGAGGGCGCTGAGAAGAGCGGGCAACGCGCAACT	1403
QY	1593	TGTTGCGGAGCGCGCCGACGCCCTGAACGCCCTTGAAGAGCTGTACGAACACCGCGCGC	1652
Db	1404	TGTTGCGGAGCGCGCCGACGCCCTGAACGCCCTTGAAGAGCTGTACGAACACCGCGCGC	1463
QY	1653	AGGCGCGTACGACGGAGCCCGTTGGCAGGAAGCACCTCCGGAACCAACAGGACGCGCTGAC	1712
Db	1464	AGGCGCGTACGACGGAGCCCGTTGGCAGGAAGCACCTCCGGAACCAACAGGACGCGCTGAC	1523
QY	1713	GCTCCGGCAGCAAGGGGGCGGAAGAGCGGCTTGCCGAACCTGAAGCCGCCGAAGCCCCGAA	1772
Db	1524	GCTCCGGCAGCAAGGGGGCGGAAGAGCGGCTTGCCGAACCTGAAGCCGCCGAAGCCCCGAA	1583
QY	1773	GCTTCCCCCTTGACCAATGTGTCCCGGAAGACGCCGACGCTGACCCGACCGGCTTAACTC	1832
Db	1584	GCTTCCCCCTTGACCAATGTGTCCCGGAAGACGCCGACGCTGACCCGACCGGCTTAACTC	1643
QY	1833	GTGGTGGGGGGCGCGCTCAGTAGACGACAAGCGCGTGTTCGTGGGCTCTTCGTAGCAA	1892
Db	1644	GTGGTGGGGGGCGCGCTCAGTAGACGACAAGCGCATGTTCGTGGGCTCTTCGTAGCAA	1703
QY	1893	GATCGTTGTACAGAAAGTCGACTACGGGGCAGGGGGCAGGGGAACGCCCATCGAAGCCGCC	1952
Db	1704	GATCGTTGTACAGAAAGTCGACTACGGGGCAGGGGGCAGGGGAACGCCCATCGAAGCCGCC	1763
QY	1953	TTTCGATCACGTGGGGCGAAGCCCGCCGACCGACGACGACGAAAGACGACGCCAGAGCGGCAC	2012
Db	1764	TTTCGATCACGTGGGGCGAAGCCCGCCGACCGACGACGACGAAAGACGACGCCAGAGCGGCAC	1823
QY	2013	GGAAGACGTAGCGGCGTAG 2031	
Db	1824	GCAAGACGTAGCGGCGTAG 1842	

	RESULT 8
AAH74881	
ID	AAH74881 standard; DNA; 1839 BP.
XX	
AC	AAH74881;
XX	
DT	29-OCT-2001 (first entry)
XX	
DE	Nucleotide sequence of altered r

[illegible]

QY 333 AGTCGAGCGCGACGGGGCGGTTTCAGGTTCTCGGGCATTTTCAGCGAAGCGCGGCAC 392  
 Db 144 AGTCGAGCGCGACGGGGCGGTTTCAGGTTCTCGGGCATTTTCAGCGAAGCGCGGCAC 203  
 QY 393 GTCGGCGTTTCGGGACGGCGGAGCGCCCGGAGTTTGAACGCATCCGTAACGAATGCCCGC 452  
 Db 204 GTCGGCGTTTCGGGACGGCGGAGCGCCCGGAGTTTGAACGCATCCGTAACGAATGCCCGC 263  
 QY 453 CGGGCGGCTCAACATGATCATTTGTCTATGAGTGTGCGGCTTCTCGCCCTGAAGGTCTAT 512  
 Db 264 CGGGCGGCTCAACATGATCATTTGTCTATGAGTGTGCGGCTTCTCGCCCTGAAGGTCTAT 323  
 QY 513 GGACGCGATTCGATTTGTCTCGGAATTTGCTCGCCCTGGGGCTGACGATTTGTTCCACTCA 572  
 Db 324 GGACGCGATTCGATTTGTCTCGGAATTTGCTCGCCCTGGGGCTGACGATTTGTTCCACTCA 383  
 QY 573 GGAAGCGCTTCTCGGCGAGGGAAACGTCATGGAACCTGATTCACCTGATTTATGCGGCTCGA 632  
 Db 384 GGAAGCGCTTCTCGGCGAGGGAAACGTCATGGAACCTGATTCACCTGATTTATGCGGCTCGA 443  
 QY 633 CGCGTCGCACAAAGATCTTCGCTGAAGTCGGCGAAGATTTCTGACACAGAAACCTTCA 692  
 Db 444 CGCGTCGCACAAAGATCTTCGCTGAAGTCGGCGAAGATTTCTGACACAGAAACCTTCA 503  
 QY 693 GCGCGAATTTGGGGGGGTACGTGCGCGGAGGCGCCTTACGCGCTTCGAGCTTTGTTGGA 752  
 Db 504 GCGCGAATTTGGGGGGGTACGTGCGCGGAGGCGCCTTACGCGCTTCGAGCTTTGTTGGA 563  
 QY 753 GACGAAGGAGATCACGCGCAACGCGCAATGTCATGTCATCAACAAGCTTCGCGCA 812  
 Db 564 GACGAAGGAGATCACGCGCAACGCGCAATGTCATGTCATCAACAAGCTTCGCGCA 623  
 QY 813 CTCGACCACTCCCTTACCAGCCCTTCGAGTTCGAGCCCGACGTAATCCGGTGGTGTG 872  
 Db 624 CTCGACCACTCCCTTACCAGCCCTTCGAGTTCGAGCCCGACGTAATCCGGTGGTGTG 683  
 QY 873 GCGTGAGATCAAGACGCACAAACACCTTCCCTTCAAGCCGGGAGTCAAGCCCGCATTTCA 932  
 Db 684 GCGTGAGATCAAGACGCACAAACACCTTCCCTTCAAGCCGGGAGTCAAGCCCGCATTTCA 743  
 QY 933 CCGGGCAGCATCACGGGGCTTTGTAAAGCGCATGACGCTGACGCGTCCCGACCCGGGG 992  
 Db 744 CCGGGCAGCATCACGGGGCTTTGTAAAGCGCATGACGCTGACGCGTCCCGACCCGGGG 803  
 QY 993 CGAGACGATTTGGGAAGAAGACCGCTTCAAGCGCTGGACCCCGCAACCGTTATGCGAAT 1052  
 Db 804 CGAGACGATTTGGGAAGAAGACCGCTTCAAGCGCTGGACCCCGCAACCGTTATGCGAAT 863  
 QY 1053 CCTTCGGACCCCGCTATTTGCGGCTTCGCGCTGAGTGTATCTACAAGAAGAAGCCGGA 1112  
 Db 864 CCTTCGGACCCCGCTATTTGCGGCTTCGCGCTGAGTGTATCTACAAGAAGAAGCCGGA 923  
 QY 1113 CGGCACGCCGACACGGAAGATGAGGGTTACCGCATTCAGCGCGACCCGATCAGCTCCG 1172  
 Db 924 CGGCACGCCGACACGGAAGATGAGGGTTACCGCATTCAGCGCGACCCGATCAGCTCCG 983  
 QY 1173 GCCGGTGAAGCTTGAATGCGGACCGATCATGAGCCCGCTGAGTGTATGAGCTTCAGGC 1232  
 Db 984 GCCGGTGAAGCTTGAATGCGGACCGATCATGAGCCCGCTGAGTGTATGAGCTTCAGGC 1043  
 QY 1233 GTGTTGGACGGCAGGGGGCGGCAAGGGCTTCCCGGGGCAAGCCATTTCTGTCCGC 1292  
 Db 1044 GTGTTGGACGGCAGGGGGCGGCAAGGGCTTCCCGGGGCAAGCCATTTCTGTCCGC 1103  
 QY 1293 CATGACAAGCTGTACTGCGAGTGTGCGCGCTCATGACTTCGAAGCGCGGGGAAGAATC 1352  
 Db 1104 CATGACAAGCTGTACTGCGAGTGTGCGCGCTCATGACTTCGAAGCGCGGGGAAGAATC 1163  
 QY 1353 GATCAAGGACTTTACCGCTGCCGTGCCGGAAGGTGTGACCCGCTCCGACCTGGGCA 1412  
 Db 1164 GATCAAGGACTTTACCGCTGCCGTGCCGGAAGGTGTGACCCGCTCCGACCTGGGCA 1223

QY 1413 GCACGAAGGCACGTGCAACGTACAGATTTGGCGGCACTGACAGATTTGTTGCGGAACCAT 1472  
 Db 1224 GCACGAAGGCACGTGCAACGTACAGATTTGGCGGCACTGACAGATTTGTTGCGGAACCAT 1283  
 QY 1473 CTTCACAAGATCAGGACGCGCGGAAGGCGACGAAGAGACGTTGGCGCTTCTGTGGAAAC 1532  
 Db 1284 CTTCACAAGATCAGGACGCGCGGAAGGCGACGAAGAGACGTTGGCGCTTCTGTGGAAAC 1343  
 QY 1533 CGCCCGACGCTTCGGCAAGCTCACTGAGGCGCCTGAGAAAGCGGGCAACGCGGAACCT 1592  
 Db 1344 CGCCCGACGCTTCGGCAAGCTCACTGAGGCGCCTGAGAAAGCGGGCAACGCGGAACCT 1403  
 QY 1593 TGTTCGGAGCGCGCGGACGCGCTGAAACGCGCTTGAAGACTGTACGAAGACCGCGCGC 1652  
 Db 1404 TGTTCGGAGCGCGCGGACGCGCTGAAACGCGCTTGAAGACTGTACGAAGACCGCGCGC 1463  
 QY 1653 AGGCGGTAAGCAGGACCGCGCTTGGCAGGAAGCACTTCGGAAGCAACAGGACCGCTGAC 1712  
 Db 1464 AGGCGGTAAGCAGGACCGCGCTTGGCAGGAAGCACTTCGGAAGCAACAGGACCGCTGAC 1523  
 QY 1713 GCTCCGGCAGCAAGGGGCGGAAGAGCGGCTTGGCGAATTTGAAGCCGCGGAACCCGAA 1772  
 Db 1524 GCTCCGGCAGCAAGGGGCGGAAGAGCGGCTTGGCGAATTTGAAGCCGCGGAACCCGAA 1583  
 QY 1773 GCTTCCCTTGACCAATGTTCCCGGAAGAGCGCGGCTGACCGCGGCGCTTAAGTC 1832  
 Db 1584 GCTTCCCTTGACCAATGTTCCCGGAAGAGCGCGGCTGACCGCGGCGCTTAAGTC 1643  
 QY 1833 GTGTGGGGCGCGCGCTCAGTAGACGACAGAGCGGCTTCTCGGGCTTCTGTAAGCA 1892  
 Db 1644 GTGTGGGGCGCGCGCTCAGTAGACGACAGAGCGGCTTCTCGGGCTTCTGTAAGCA 1703  
 QY 1893 GATCGTTGTACGAAGTCACTACGGGCGAGGGGCGAAGCGGCCCATTCAGAAAGCGGCG 1952  
 Db 1704 GATCGTTGTACGAAGTCACTACGGGCGAGGGGCGAAGCGGCCCATTCAGAAAGCGGCG 1763  
 QY 1953 TTGATCACGTGGGCGAAGCGCGGACCGGACGACGACGAGAGCGGCCGAGGAGCGGCAC 2012  
 Db 1764 TTGATCACGTGGGCGAAGCGCGGACCGGACGACGAGAGCGGCCGAGGAGCGGCAC 1823  
 QY 2013 GGAAGCGTAGCGCG 2028  
 Db 1824 GGAAGCGTAGCGCG 1839

RESULT 9  
 AAH74882  
 ID AAH74882 standard; DNA; 1833 BP.  
 XX  
 AC AAH74882;  
 XX  
 DT 29-OCT-2001 (first entry)  
 XX  
 DE Nucleotide sequence of altered recombinase 1C1.  
 XX  
 KW Recombinase; genetic engineering; monogenic disorder; ADA deficiency;  
 KW cystic fibrosis; familial-hypercholesterolaemia; haemophilia;  
 KW chronic granulomatous disease; Duchenne's muscular dystrophy;  
 KW Fanconi's disease; anemia; sickle-cell anemia; Gaucher's disease;  
 KW Hunter's syndrome; X-linked severe combined immunodeficiency; SCID;  
 KW infectious disease; acquired disorder; tumour; cancer; ss.  
 XX  
 OS Synthetic.  
 OS Bacteriophage.  
 XX  
 FH Key  
 FT CDS Location/Qualifiers  
 FT 1..1833  
 FT /tag= a  
 FT /product= "altered recombinase 1C1"  
 PN WO200161049-A1.  
 XX 23-AUG-2001.  
 PD



XX 16-FEB-2001; 2001WO-US05269.  
XX  
XX 18-FEB-2000; 2000US-0183759.  
XX  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
XX  
PI Calos MP, Scilmentl CR;  
XX  
DR WPI; 2001-522610/57.  
DR P-PSDB; AAG63849.  
XX  
XX Identifying altered recombinase. Involves hybridizing recombination  
PT sites in cells having specific coding sequence, transformed with  
PT altered recombinase gene, and isolating cells having product of the  
PT sequence -  
XX  
PS Disclosure; Fig 10; 101pp; English.  
XX  
XX The specification describes a method for identifying altered  
CC recombinases. The method comprises transforming cells with a first  
CC plasmid comprising two recombination sites and a coding sequence of  
CC interest, and a second plasmid encoding an altered recombinase; allowing  
CC recombination of the two recombination sites using the altered  
CC recombinase; screening and isolating transformed cells comprising the  
CC product of the sequence of interest; and identifying the altered  
CC recombinase. The altered recombinase is useful for site-specifically  
CC integrating a polynucleotide sequence of interest in a genome of a cell.  
CC It is also useful in genetic engineering of chromosomes of higher cells,  
CC and for the generation of transgenic cells, tissues, plants and animals.  
CC The altered recombinase is useful for treating monogenic disorders,  
CC e.g. ADA deficiency, cystic fibrosis, familial-hypercholesterolaemia,  
CC anemia, chronic granulomatous disease, Duchenne's muscular dystrophy,  
CC haemophilia, Fanconi's disease, sickle-cell anemia, Gaucher's disease,  
CC Hunter's syndrome and X-linked severe combined immunodeficiency (SCID),  
CC infectious diseases including viral and bacterial infections, acquired  
CC disorders including solid tumours and haematopoietic cancers such as  
CC leukaemias and lymphomas, and other cancers. The present sequence  
CC encodes an altered recombinase of the invention.  
XX  
SQ Sequence 1833 BP; 387 A; 544 C; 611 G; 291 T; 0 other;

Query Match 87.8%; Score 1783; DB 22; Length 1833;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1800; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 213 GATGTGACGAGTACGGGGTGTCTACGACCGTCACTCGCGGAGCGGAGATTCGAGCGC 272  
DB 24 GGTGACACGACGCGGGTCTTACGACCGTCACTCGCGGAGCGGAGATTCGAGCGC 83  
QY 273 AGCAAGCCACGACGACAGCGGTAGCGCCCAACGAGACAGAGCGCGCCGACCTTCAGCGCA 332  
DB 84 AGCAAGCCACGACGACAGCGGTAGCGCCCAACGAGACAGAGCGCGCCGACCTTCAGCGCA 143  
QY 333 AGTCGAGCGGAGCGGGGGCGGTTCAAGTTCGTCGGGCATTTCAAGCGCGCGGCAC 392  
DB 144 AGTCGAGCGGAGCGGGGGCGGTTCAAGTTCGTCGGGCATTTCAAGCGCGCGGCAC 203  
QY 393 GTCGGCGTTGGGAGCGGCGGAGCGCGGAGTTGCAAGCATCTGAACGAATGCCGCGC 452  
DB 204 GTCGGCGTTGGGAGCGGCGGAGCGCGGAGTTGCAAGCATCTGAACGAATGCCGCGC 263  
QY 453 CGGCGGCTCAACATGATCATTTGTATGACGTGTGCGGCTTCTCGCGCTGAAGTCAT 512  
DB 264 CGGCGGCTCAACATGATCATTTGTATGACGTGTGCGGCTTCTCGCGCTGAAGTCAT 323  
QY 513 GGACGCAATTCGATGTCTCGAATTCGTCGCCCTGGCGGTGACGATTTTCCACTCA 572  
DB 324 GGACGCAATTCGATGTCTCGAATTCGTCGCCCTGGCGGTGACGATTTTCCACTCA 383  
QY 573 GGAAGCGCTCTTCGGGAGGGAAGCTGACCTGATTCACCTGATATGCGGCTCGA 632  
DB 384 GGAAGCGCTCTTCGGGAGGGAAGCTGACCTGATTCACCTGATATGCGGCTCGA 443

QY 633 CGCGTCGACAAAGAAATCTTCGCTGAAGTCGGCGAAGATTCGACAGGAACCTTCA 692  
DB 444 CGCGTCGACAAAGAAATCTTCGCTGAAGTCGGCGAAGATTCGACAGGAACCTTCA 503  
QY 693 GCGGCAATTCGGCGGTACGTCGGCGGGAAGCGCCCTTACGGCTTCGAGCTTGTTCGGA 752  
DB 504 GCGGCAATTCGGCGGTACGTCGGCGGGAAGCGCCCTTACGGCTTCGAGCTTGTTCGGA 563  
QY 753 GACGAAGGAGATCAGCGCCCAACGGCCGATGTCATGTCTCATCAACAGCTTGGCGA 812  
DB 564 GACGAAGGAGATCAGCGCCCAACGGCCGATGTCATGTCTCATCAACAGCTTGGCGA 623  
QY 813 CTCGACCACTCCCTTACCGGACCTTCGAGCTTCGAGCCCGACGTAATCCGCTGTGTG 872  
DB 624 CTCGACCACTCCCTTACCGGACCTTCGAGCTTCGAGCCCGACGTAATCCGCTGTGTG 683  
QY 873 GCGTGATCAAGACGACAAACACCTTCCCTTCAAGCCGGGAGTCAGCCGCATTTCA 932  
DB 684 GCGTGATCAAGACGACAAACACCTTCCCTTCAAGCCGGGAGTCAGCCGCATTTCA 743  
QY 933 CCCGGGACGATCAGCGGGGCTTGTAAAGCGCATGACGCTGACCGCGTCCGACCCGGG 992  
DB 744 CCCGGGACGATCAGCGGGGCTTGTAAAGCGCATGACGCTGACCGCGTCCGACCCGGG 803  
QY 993 CGAGACGATTGGGAGAGAACCCGCTTCAAGCGCTGGGACCCGGCAACCGTTATGCGAAT 1052  
DB 804 CGAGACGATTGGGAGAGAACCCGCTTCAAGCGCTGGGACCCGGCAACCGTTATGCGAAT 863  
QY 1053 CCTTCGGGACCCGCTATTTGCGGGCTTCGCGCTGAGGTGATCTCAAGAAAGACCGGA 1112  
DB 864 CCTTCGGGACCCGCTATTTGCGGGCTTCGCGCTGAGGTGATCTCAAGAAAGACCGGA 923  
QY 1113 CGGACAGCGGACGACGAAGATTGAGGGTTACCGCATTCAGCGCGGACCCGATCAGCTCCG 1172  
DB 924 CGGACAGCGGACGACGAAGATTGAGGGTTACCGCATTCAGCGCGGACCCGATCAGCTCCG 983  
QY 1173 GCCGTCGAGCTTGTATGCGGACCGCATTCAGACCCGCTGAGTGTATGAGCTTCAGGC 1232  
DB 984 GCCGTCGAGCTTGTATGCGGACCGCATTCAGACCCGCTGAGTGTATGAGCTTCAGGC 1043  
QY 1233 GTGCTTGACGCGGAGGGGGCGGCGCAAGGGCTTCCCGGGGCAAGCCATTCTGTCCGC 1292  
DB 1044 GTGCTTGACGCGGAGGGGGCGGCGCAAGGGCTTCCCGGGGCAAGCCATTCTGTCCGC 1103  
QY 1293 CATGACAAGCTGTACTGCGAGTGTGGCGCTCATGACTTCGAAGCGCGGGAAGAATC 1352  
DB 1104 CATGACAAGCTGTACTGCGAGTGTGGCGCTCATGACTTCGAAGCGCGGGAAGAATC 1163  
QY 1353 GATCAAGACTCTTACCGCTGCGCGTCCGGAAGGTGTGACCCGCTCCGACCTGGGCA 1412  
DB 1164 GATCAAGACTCTTACCGCTGCGCGTCCGGAAGGTGTGACCCGCTCCGACCTGGGCA 1223  
QY 1413 GCACGAAGGACGTCACAGTCAAGTGGGCGACTCGACAAAGTTCTGTGGAAGCAT 1472  
DB 1224 GCACGAAGGACGTCACAGTCAAGTGGGCGACTCGACAAAGTTCTGTGGAAGCAT 1283  
QY 1473 CTTCAACAAGATCAGGCAAGCGCGCAAGGCGCAAGAGACGTTGGCGCTTCTGTGGAAGC 1532  
DB 1284 CTTCAACAAGATCAGGCAAGCGCGCAAGGCGCAAGAGACGTTGGCGCTTCTGTGGAAGC 1343  
QY 1533 CGCCGAGCGCTTGGCAAGCTCACTGAGCGCTGAGAAGAGCGGCAAGCGGGAACCT 1592  
DB 1344 CGCCGAGCGCTTGGCAAGCTCACTGAGCGCTGAGAAGAGCGGCAAGCGGGAACCT 1403  
QY 1593 TGTTCGAGGCGCGCGACGCGCCCTGAAGCGCCCTGAAGAGCTGTACGAAGACCGCGGC 1652  
DB 1404 TGTTCGAGGCGCGCGACGCGCCCTGAAGCGCCCTGAAGAGCTGTACGAAGACCGCGGC 1463  
QY 1653 AGGCGGTACGAGGAGCCCGTTGGAGAGGACACTTCCGGAAGCAAGGACGCGCTGAC 1712  
DB 1464 AGGCGGTACGAGGAGCCCGTTGGAGAGGACACTTCCGGAAGCAAGGACGCGCTGAC 1523

QY	1713	GCTCCGGCAGCAAGGGCGGGAAGACGGCTTGCCGAATTGAAGCCGCCGAAGCCCCGAA	1772
Db	1524	GCTCCGGCAGCAAGGGCGGGAAGACGGCTTGCCGAATTGAAGCCGCCGAAGCCCCGAA	1583
QY	1773	GCTTCCCCTTGACCAATGTTCCTCCGAAGACGCCGAGCTGACC CGCGCCCTAAGTC	1832
Db	1584	GCTTCCCCTTGACCAATGTTCCTCCGAAGACGCCGAGCTGACC CGCGCCCTAAGTC	1643
QY	1833	GTGTTGGGGGCGCGCTCAGTAGACGACAAGCGCTTCGTCGGGCTCTTCGTAGACAA	1892
Db	1644	GTGTTGGGGGCGCGCTCAGTAGACGACAAGCGCTTCGTCGGGCTCTTCGTAGACAA	1703
QY	1893	GATCGTTGTACAGAACTCGACTACGGGCAGGGGGCAGGGAACGCCCATCGAGAAGCGGC	1952
Db	1704	GATTTGTTGTACAGAACTCGACTACGGGCAGGGGGCAGGGAACGCCCATCGAGAAGCGGC	1763
QY	1953	TTCGATCACGTGGGGCCAAGCCGCCGACCGACGACGACGAAGACGACGCCCCAGACGGCAC	2012
Db	1764	TTCGATCACGTGGGGCCAAGCCGCCGACCGACGACGACGAAGACGACG-CCAGGACGGGCAC	1822
QY	2013	GGAAGACGTAG	2023
Db	1823	GGAAGACGTAG	1833

RESULT 10  
AAD19829  
ID AAD19829 standard; DNA; 912 BP.

AAD19829;

18-DEC-2001 (first entry)

Green fluorescent protein-n-1, syngFP1 gene with ST-LS1 intron sequence.

Cestrum yellow leaf curling virus; CMYLCV; transcription; transgenic plant; green fluorescent protein 1; GFP1; ds.

Solanum tuberosum.  
Unidentified.

Key	Location/Qualifiers
Intron	278..465

```
/*tag- a  
/note- "Solanium tuberosum ST-LS1 intron"
```

WO200173087-A1.

04-OCT-2001.

26-MAR-2001; 2001WO-EP03408.

27-MAR-2000; 2000GB-0007427.

26-JAN-2001; 2001EP-0101802.

(SYGN ) SYNGENTA PARTICIPATIONS AG.

Hohn T, Stavolone L, De Haan PT, Ligon HT, Kononova M;

WPI; 2001-616524/71.

PT Novel DNA sequence obtained from genome of Cestrum yellow leaf curling  
PT virus for conferring constitutive expression of an associated desired  
PT polynucleotide -

Example 19; Page 66; 100pp; English.

CC The invention relates to Cestrum yellow leaf curling virus (CmYLCV) novel  
CC DNA sequences which functions as transcription promoters of an associated  
CC polynucleotide sequence. These CmYLCV DNA molecules confers constitutive  
CC expression of associated polynucleotide sequences. The invention also  
CC relates to recombinant DNA sequences containing promoter sequences which

CC are used for creating transgenic plants expressing DNA of interest at all  
CC times and in most tissues and organs. The present DNA sequence is green  
CC fluorescent protein-h1, syngFP1 gene with ST-LS1 intron sequence. Plant  
CC optimised GFP reporter gene is used in the construction of plant  
CC transformation vectors.

SQ Sequence 912 BP; 245 A; 245 C; 240 G; 182 T; 0 other;  
XX

Query Match	9.8%;	Score 198.6;	DB 22;	Length 912;
Best Local Similarity	95.8%;	Pred. No. 1e-34;		
Matches 204;	Conservative	0;	Mismatches 9;	Indels 0;
				Gaps 0

QY    21 CCGGGTGGATACGTAAGTTCTGCCTCTACCTTGATATATATAAATTAATTCATTAA 80  
     | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    265 CCGAGAGGCCATCAGTAAGTTCTGCCTCTACCTTGATATATATAAATAATTCATTAA 324

OY	81	TTAGTAGTAATATAATATTTCAAAATATTTTTTCAAATAAAGAATGTAGTATATACGA	140
Db	325	TTAGTAGTAATATAATATTTCAAATATTTTTTCAAATAAAGAATGTAGTATATACGA	384

QY 141 ATGCTTTCTGTAGTTTATAAGTGTATATTTTAATTTAACTTTTCTAATATATGA 200  
 |||||  
 Db 385 ATGCTTTTCTGTAGTTTATAAGTGTATATTTTAATTTAACTTTTCTAATATATGA 444

Oy	201	CCAAATTTGTTGATGTGCAGGTACGGCGGTGC	233
Db	445	CCAAATTTGTTGATGTGCAGGTGCAGGAGCGC	477

RESULT 11  
AAD19834

ID AAD19834 standard; DNA; 1577 BP.

AC AAD19834;

DT	18-DEC-2001	(first entry)
----	-------------	---------------

Promoter-reporter cassette #3 to construct plant transformation vector

KW Cestrum yellow leaf curling virus; CMYCV; transcription;  
KW transgenic plant; nopaline synthase; NOS; green fluorescent protein-1;  
KW GFP1; ds.

OS	Chimeric - Cestrum yellow leaf curling virus.
OS	Chimeric - Agrobacterium tumefaciens.
OS	Chimeric - Unidentified.

	Key	Location/Qualifiers
FH	promoter	1..354
FT		

FT	misc feature	FT
/note= "CMYLCV cmc promoter fragment"		
380..1292		

FT	3	1304..1577	Corresponds to SyngFP1 gene"
FT	terminator		

```
FT      /*tag= c
FT      /note= "Nopaline synthase (NOS) terminator"
```

PN WO200173087-A1.

PD 04-OCT-2001.

PF 26-MAR-2001; 2001WO-EP03408.

PR 27-MAR-2000; 2000GB-0007427.

PR 26-JAN-2001; 2001EP-0101802.

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

PI Hohn T, Stavolone L, De Haan PT, Ligon HT, Kononova M;

DR WPI; 2001-616524/71.



```
FH Key Location/Qualifiers
FT promoter 1..1332
FT /tag= a
FT /note= "Arabidopsis ubiquitin 3 promoter, Ubq3(At)"
FT misc_feature 1738..2658
FT /tag= b
FT /note= "Corresponds to green fluorescent protein (GFP)
FT gene"
FT terminator 2670..2943
FT /tag= c
FT /note= "Nopaline synthase (NOS) terminator"
XX
XX WO200173087-A1.
XX
XX PD 04-OCT-2001.
XX
XX PF 26-MAR-2001; 2001WO-EP03408.
XX
XX PR 27-MAR-2000; 2000GB-0007427.
XX 28-APR-2000; 2000GB-0010486.
XX 26-JAN-2001; 2001EP-0101802.
XX
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX PI Hohn T, Stavolone L, De Haan PT, Ligon HT, Kononova M;
XX
XX DR WPI; 2001-616524/71.
XX
XX PT Novel DNA sequence obtained from genome of Cestrum yellow leaf curling
XX virus for conferring constitutive expression of an associated desired
XX polynucleotide
XX
XX PS Example 19; Page 93-94; 100pp; English.
XX
XX CC The invention relates to Cestrum yellow leaf curling virus (CmYLCV) novel
XX CC DNA sequences which functions as transcription promoters of an associated
XX CC polynucleotide sequence. These CmYLCV DNA molecules confers constitutive
XX CC expression of associated polynucleotide sequences. The invention also
XX CC relates to recombinant DNA sequences containing promoter sequences which
XX CC are used for creating transgenic plants expressing DNA of interest at all
XX CC times and in most tissues and organs. The present DNA sequence is a
XX CC Ubq3(At)-syngFP1-NOS terminator expression cassette which comprises
XX CC Arabidopsis thaliana ubiquitin 3 promoter, Ubq3(At) operatively linked
XX CC with green fluorescent protein-1, syngFP1 reporter gene and
XX CC Agrobacterium tumefaciens nopaline synthase (NOS) terminator. This
XX CC expression cassette is used for the construction of plant transformation
XX CC vectors.
XX
XX SQ Sequence 2943 BP; 886 A; 544 C; 584 G; 929 T; 0 other;

Query Match 9.8%; Score 198.6; DB 22; Length 2943;
Best Local Similarity 95.8%; Pred. No. 1.6e-34;
Matches 204; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 21 CCGGGTGATACGTAAAGTTCTGCTTCTACCTTGATATATATAATATATCATTA 80
DB 2011 CCGAGGCGCTACGTAAAGTTCTGCTTCTACCTTGATATATATAATATATCA 2070
OY 81 TTAGTAGTAATATAATATTTTCAAAATATAAAGAAATGTAGTATATAGCA 140
DB 2071 TTAGTAGTAATATAATATTTTCAAAATATAAAGAAATGTAGTATATAGCA 2130
OY 141 ATGCTTTCTGTAGTTTATAAGTGTATATTTTAATTATTAACCTTTCTAATATATGA 200
DB 2131 ATGCTTTCTGTAGTTTATAAGTGTATATTTTAATTATTAACCTTTCTAATATATGA 2190
OY 201 CCAAAATTTGTGATGTGACAGGACGGGGTGC 233
DB 2191 CCAAAATTTGTGATGTGACAGGACGGCGC 2223

RESULT 14
AAD19830
```

```
ID AAD19830 standard; DNA; 2001 BP.
XX
XX AC AAD19830;
XX
XX DT 18-DEC-2001 (first entry)
XX
XX DE Beta-glucuronidase (GUS) reporter gene with ST-LS1 intron sequence.
XX
XX KW Cestrum yellow leaf curling virus; CmYLCV; transcription;
XX transgenic plant; beta-glucuronidase; GUS; ds.
XX
XX OS Solanum tuberosum.
XX OS unidentified.
XX
XX FH Key Location/Qualifiers
FT intron 385..576
FT /tag= a
FT /note= "Solanum tuberosum ST-LS1 intron"
XX
XX PN WO200173087-A1.
XX
XX PD 04-OCT-2001.
XX
XX PF 26-MAR-2001; 2001WO-EP03408.
XX
XX PR 27-MAR-2000; 2000GB-0007427.
XX 28-APR-2000; 2000GB-0010486.
XX 26-JAN-2001; 2001EP-0101802.
XX
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX PI Hohn T, Stavolone L, De Haan PT, Ligon HT, Kononova M;
XX
XX DR WPI; 2001-616524/71.
XX
XX PT Novel DNA sequence obtained from genome of Cestrum yellow leaf curling
XX virus for conferring constitutive expression of an associated desired
XX polynucleotide
XX
XX PS Example 19; Page 67-68; 100pp; English.
XX
XX CC The invention relates to Cestrum yellow leaf curling virus (CmYLCV) novel
XX CC DNA sequences which functions as transcription promoters of an associated
XX CC polynucleotide sequence. These CmYLCV DNA molecules confers constitutive
XX CC expression of associated polynucleotide sequences. The invention also
XX CC relates to recombinant DNA sequences containing promoter sequences which
XX CC are used for creating transgenic plants expressing DNA of interest at all
XX CC times and in most tissues and organs. The present DNA sequence is
XX CC beta-glucuronidase (GUS) reporter gene with ST-LS1 intron sequence.
XX CC The GIG (GUS reporter gene with intron) gene is used in the construction
XX CC of plant transformation vectors.
XX
XX SQ Sequence 2001 BP; 532 A; 451 C; 535 G; 483 T; 0 other;

Query Match 9.6%; Score 195; DB 22; Length 2001;
Best Local Similarity 100.0%; Pred. No. 9e-34;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 TACGTAGTTTCTGCTTCTACCTTTGATATATATAATAATTATCATTAATAGTAGTA 89
DB 383 TACGTAGTTTCTGCTTCTACCTTTGATATATATAATAATTATCATTAATAGTAGTA 442
OY 90 ATATAATATTTCAAAATATTTTTCAAAATAAAGAAATAGTATATAGCAATTGCTTTT 149
DB 443 ATATAATATTTCAAAATATTTTTCAAAATAAAGAAATAGTATATAGCAATTGCTTTT 502
OY 150 CTGTAGTTTATAAGTGTGATATTTTAAATTATTAACCTTTCTAATATATAGCAAAATTT 209
DB 503 CTGTAGTTTATAAGTGTGATATTTTAAATTATTAACCTTTCTAATATATAGCAAAATTT 562
OY 210 GTTGATGTGACAGTA 224
DB 563 GTTGATGTGACAGTA 577
```

RESULT 15  
ABK15667  
ID ABK15667 standard; DNA; 2001 BP.  
XX  
AC ABK15667;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Glutathione-S-transferase with S. tuberosum intron, GIG, sequence.  
XX  
KW ds; lipoxigenase; RCI-1; transgenic; plant antifungal; pNOV2347;  
KW rice chemically induced cDNA; promoter; transit peptide; plastid;  
KW fungal mycotoxin inhibitor; plant breeding; glutathione-S-transferase;  
KW GUS; GIG.  
XX  
OS Chimeric - Escherichia coli.  
OS Chimeric - Solanum tuberosum.  
XX  
PN WO200206490-A1.  
XX  
PD 24-JAN-2002.  
XX  
PF 12-JUL-2001; 2001WO-EP08085.  
XX  
PR 13-JUL-2000; 2000GB-0017275.  
PR 15-SEP-2000; 2000GB-0022739.  
XX  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
PA (UYZU-) UNIV ZUERICH.  
XX  
PI Dudler R, Schaffrath , Lawton KA;  
XX  
DR WPI; 2002-188550/24.  
XX  
PT Novel isolated nucleic acid encoding a promoter which is capable of  
PT driving chemically inducible but not wound- or pathogen-inducible  
PT expression of an associated nucleotide sequence  
XX  
PS Example 17; Page 75; 88pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (a promoter of  
CC rice chemically induced cDNA (RCI-1), which encodes a lipoxigenase)  
CC capable of driving chemically-inducible but not wound- or pathogen-  
CC inducible expression of an associated nucleotide sequence. Also  
CC included are the RCI-1 cDNA, its encoded protein, a 4.5kb genomic clone  
CC for the lipoxigenase gene, promoter fragments, the lipoxigenase transit  
CC peptide which directs expressed proteins to the plastid, a vector  
CC comprising the promoter or fragments and a transgenic plant comprising  
CC the vector. The promoter or fragments are useful for expressing a  
CC nucleotide sequence of interest. The transit peptide is useful for  
CC targeting an associated protein of interest to plastids. A nucleic acid  
CC which expresses polypeptide having lipoxigenase activity is useful for  
CC inhibiting fungal mycotoxins when transformed into a plant. The  
CC lipoxigenase is useful for inhibiting fungal mycotoxins. The promoter is  
CC useful for regulating transcription of a chemically inducible but not  
CC wound or pathogen inducible gene, which involves applying a chemical  
CC regulator to a plant or seed containing a chemically regulatable  
CC nucleotide sequence. Transgenic plants as described above are useful for  
CC breeding improved plant lines that for example increase the effectiveness  
CC of conventional methods such as herbicide or pesticide treatment or allow  
CC to dispense with the methods due to their modified genetic properties.  
CC New crops with improved stress tolerance can be obtained that, due to  
CC their optimised genetic equipment yield harvested product of better  
CC quality than products that were not able to tolerate comparable adverse  
CC developmental conditions. The present sequence is the glutathione-S-  
CC transferase (GUS) sequence (containing an intron from solanum  
CC tuberosum, referred to as GIG) whose expression is driven from rice  
CC lipoxigenase promoter regions in a vector construct.  
XX  
SQ Sequence 2001 BP; 532 A; 451 C; 535 G; 483 T; 0 other;

Query Match 9.6%; Score 195; DB 24; Length 2001;  
Best Local Similarity 100.0%; Pred. No. 9e-34;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 30 TACGTAAGTTTCTGCTTCTACCTTTGATATATATATAATTAATTAATTAGTAGTA 89  
DB 383 TACGTAAGTTTCTGCTTCTACCTTTGATATATATATAATTAATTAATTAGTAGTA 442  
QY 90 ATATAATATTTCAATATATTTTTCAAATAAAGAATGTAGTATATAGCAATTCCTTT 149  
DB 443 ATATAATATTTCAATATATTTTTCAAATAAAGAATGTAGTATATAGCAATTCCTTT 502  
QY 150 CTGTAGTTTATAAGTGTGTATATTTTAAATTTTAACTTTTCTAATATATGACCAAAATTT 209  
DB 503 CTGTAGTTTATAAGTGTGTATATTTTAAATTTTAACTTTTCTAATATATGACCAAAATTT 562  
QY 210 GTTGATGTGCAGGTA 224  
DB 563 GTTGATGTGCAGGTA 577

Search completed: May 11, 2003, 18:05:03  
Job time : 536.5 secs





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 17:39:39 ; Search time 3307.5 Seconds  
(without alignments)  
9944.984 Million cell updates/sec

Title: US-09-940-550A-9  
Perfect score: 2031  
Sequence: 1 atgycacaagggtgtgtgac.....cggaaagacgtagcgcgtag 2031

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	125.8	6.2	675	13	BM109636 EST557172
C 2	125.8	6.2	804	14	BQ512568 EST619983
C 3	125.8	6.2	847	14	BQ512569 EST619984
C 4	60.4	3.0	621	17	CNS04POX AL301578 Tetraodon
C 5	58	2.9	928	17	CNS00DKY AL071865 Drosophila
C 6	58	2.9	964	17	CNS017VX AL108567 Drosophila

C 7	57.4	2.8	752	17	BH585665	BH585665 BOGSL20TF
C 8	57.4	2.8	986	17	CNS0170T	AL107447 Drosophila
C 9	56.4	2.8	1101	17	CNS0175Y	AL108460 Drosophila
C 10	56	2.8	780	17	CNS00T74	AL089326 Arabidops
C 11	56	2.8	1201	17	CNS01038	AL098462 Drosophila
C 12	55.8	2.7	1101	17	CNS00ZAL	AL097431 Drosophila
C 13	55.6	2.7	1190	17	CNS020N7	AL206908 Tetraodon
C 14	55.4	2.7	1101	17	CNS003BD	AL064091 Drosophila
C 15	55.2	2.7	975	17	BH179465	BH179465 014_P10-
C 16	55	2.7	644	17	AQ961004	AQ961004 LERFI58TF
C 17	55	2.7	720	17	CNS00XZ8	AL095522 Arabidops
C 18	55	2.7	925	17	CNS0091P	AL053013 Drosophila
C 19	54.8	2.7	859	12	BF272342	BF272342 GA_Eb001
C 20	54.2	2.7	1101	17	CNS00EVL	AL069706 Drosophila
C 21	54.2	2.7	1101	17	CNS00EVL	AL069706 Drosophila
C 22	54	2.7	920	17	CNS01UQL	AL168150 Tetraodon
C 23	53.8	2.6	602	17	B28108	B28108 T2G12TRD TA
C 24	53.8	2.6	1101	17	CNS00CYH	AL060100 Drosophila
C 25	53.8	2.6	1101	17	CNS002DB	AL062360 Drosophila
C 26	53.4	2.6	454	17	B96681	B96681 T3001TR TAM
C 27	53.4	2.6	748	17	AQ286687	AQ286687 RPT11-81
C 28	53.4	2.6	876	17	CNS009G1	AL053529 Drosophila
C 29	53.2	2.6	272	14	BQ596436	BQ596436 PfESToab3
C 30	53.2	2.6	423	17	CNS00S62	AL087992 Arabidops
C 31	53.2	2.6	693	14	BQ838839	BQ838839 WHE3593_C
C 32	53.2	2.6	807	17	BH438258	BH438258 BOGLZ67TF
C 33	53.2	2.6	812	17	BH178455	BH178455 011_J_02-
C 34	53.2	2.6	812	17	CNS07KRM	AL615412 T3 end of
C 35	53.2	2.6	1101	17	CNS016LI	AL106896 Drosophila
C 36	53	2.6	364	17	AQ583787	AQ583787 RPT1-11-4
C 37	53	2.6	1101	17	CNS00DT7	AL075293 Drosophila
C 38	52.6	2.6	1101	17	CNS00EPO	AL069493 Drosophila
C 39	52.4	2.6	673	17	CNS06UWL	AL416395 T7 end of
C 40	52.4	2.6	905	17	CNS00KHX	AL077798 Drosophila
C 41	52.2	2.6	863	12	BF264892	BF264892 HY_CEA001
C 42	52.2	2.6	1101	17	CNS00EQL	AL069526 Drosophila
C 43	52	2.6	996	17	CNS00FUH	AL071063 Drosophila
C 44	52	2.6	1043	17	CNS0145P	AL103735 Drosophila
C 45	52	2.6	1292	13	BM463105	BM463105 AGENCOURT

ALIGNMENTS

RESULT 1  
BM109636/c 675 bp mRNA linear EST 26-NOV-2001  
LOCUS EST557172 potato roots Solanum tuberosum cDNA clone cPR04H6 5' end,  
DEFINITION mRNA sequence.

ACCESSION BM109636  
VERSION BM109636.1 GI:17070714  
KEYWORDS EST.  
SOURCE potato.  
ORGANISM Solanum tuberosum

REFERENCE 1 (bases 1 to 675)  
AUTHORS van der Hoeven,R., Sun,H., Karamycheva,S.A., Tsai,J., Van Aken,S.,  
Utterback,T., Chiemiango,A., Bougri,O., Buell,C.R., Ronning,C.,  
Tanksley,S. and Baker,B.

TITLE Generation of ESTs from potato roots  
JOURNAL Unpublished (2001)  
COMMENT Contact: Research Genetics, Libraries Division  
Tel: 1-800-711-6195  
Email: cdna@resgen.com

FEATURES  
source Location/Qualifiers  
1..675  
/organism="Solanum tuberosum"  
/cultivar="Kennebec"

```

/db_xref="taxon:4113"
/clone="CPRO4H6"
/clone_lib="potato roots"
/tissue_type="roots"
/dev_stage="in vitro grown stem cuttings"
/lab_host="SOLR"
/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Cornell University, Tanksley lab;
sequencing: The Institute for Genomic Research. Roots were
isolated from in vitro grown stem cuttings on CM medium.
Roots were isolated two weeks after placing the stem
cuttings from in vitro grown plants on medium."

```

Query Match	6.28;	Score 125.8;	DB 13;	Length 675;
Best Local Similarity	83.98;	Pred. No. 4.9e-18;		
Matches 167; Conservative	0;	Mismatches 27;	Indels 5;	Gaps 2;

OY      31 ACGTAGCTTCTGCTCTCTACCTTTGATATATATAAATTAATTCATTAACTTAGTAGTAA    90  
       | || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db      663 AGGTTCGTGTTACTCTCTACCTTGATATATAT--ATAATATCAATTAATTAGTAGTAA    607

QY    91 TATAATATTTCGAATATTTTTCCTCAAAATAAAGAATGTAGTATATAGCAATTGCTTTC 150  
       | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    606 AATAATATTTCCAT - TTTTTTTCACAAATAAAGAATGTAAATTAAGCAATTGCTTTC 549

QY 151 TGTAGTTTAAGTGTGTATATTTTAATTATACTTTTCRAATATATGACCAAAATTTC 210  
|||||  
Db 548 TGTAGTTTAAGTGTGTATATTTTAATTATACTTTTCRAATATATGAGCAAAACATG 489

QY	211	TTGATGTCAGGTACCGG	229
Db	488	GTGATGTTAGGGAAGG	470

RESULT 2	LOCUS	DEFINITION
BQ512568/c	BQ512568	ESTc19983
	804 bp	Generation of a set of potato CDNA clones for microarray analyses mixed potato tissues Solanum tuberosum CDNA clone STMHZ22 5' end, mRNA sequence.
	mRNA	linear EST 10-JUN-2002

ACCESSION	BQ512568	
VERSION	BQ512568.1	GI:21371437
KEYWORDS	EST.	

ORGANISM

REFERENCE  
AUTHORS  
1 (bases 1 to 804)  
Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.

TITLE	Generation of a set of potato cDNA clones for microarray analyses
JOURNAL	Unpublished (2002)
COMMENT	Other_ESTs: EST619984

The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: potato@tigr.org  
This clone is available through the Research Genetics, contact the  
Research Genetics for further information 1-800-711-6195 or  
cdna@resgen.com  
Seq primer: T3.

FEATURES	Location/Qualifiers
source	1. .804

```

1.004
/organism="Solanum tuberosum"
/cultivar="Kennebec or Binjite"
/db_xref="taxon:4113"
/clone="STMH222"
/clone_lib="Generation of a set of potato cDNA clones for
microarray analyses mixed potato tissues"

```

```

/tissue_type="mixed tissues"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants, petioles, germinating eyes
, tubers, or roots."
BASE COUNT      284 a      149 c      116 g      255 t
ORIGIN

```

Query Match	6.28;	Score 125.8;	DB 14;	Length 804;
Best Local Similarity	83.98;	Pred. No. 5.1e-18;		
Matches 167; Conservative	0;	Mismatches 27;	Indels 5;	Gaps 2;

QY    31 ACGTAGCTTTCGCTCTACCTTGATATATATAAATTATCAATTAAGTAGTAA    90  
       | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db    663 AGGTGGTTACTCTACCTTGATATAT--ATAATATCATTAATTAGTAGTAA    607

QY    91 TATAATATTCCAATATTTTTCACAAATAAAGAATGTAGTAATATGCAATTGCTTTTC 150  
       | | | | | | | | | | | | | | | | | | | | | | | |  
Db    606 AATAATATTTCAT--TTTTTTTCAAATAAAAAGATGTAAATTAAAGCAATTGCCTTTTC 549

QY 151 TGTAGTTTATAGTGTGTATATTTTAATTATACCTTTTCTAATATATGACCAAAATTTG 210  
|||||  
Db 548 TGTAGTTTATAGTGTGTATATTTTAATTATACCTTTTCTAATATATGAGCAAAACATG 489

Qy 211 TTGATGTCAGGTACGGG 229  
 ||||| ||| | ||  
 Db 488 GTGATGTTAGGGAAGGG 470

RESULT 3	LOCUS	DEFINITION
BQ512569	BQ512569	EST619984 Generation of a set of potato cDNA clones for microarray analyses mixed potato tissues Solanum tuberosum cDNA clone STMHZ22 3' end, mRNA sequence.

ACCESSION	BQ512569	
VERSION	BQ512569.1	GI:21371438
KEYWORDS	EST.	

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asterridae; euasterids I; Solanales; Solanaceae; Solanum.  
1 (bases 1 to 847)  
Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,  
REFERENCE  
AUTHORS

TITLE	Generation of a set of potato cDNA clones for microarray analyses
JOURNAL	Unpublished (2002)
COMMENT	Other_ESTs: EST619983

The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: potato@tigr.org  
This clone is available through the Research Genetics, contact the  
Research Genetics for further information 1-800-711-6195 or  
cdna@resgen.com  
Seq primer: T7.

FEATURES	Location/Qualifiers
source	1. .847

```

/organism="Solanum tuberosum"
/cultivar="Kennebec or Binjete"
/db_xref="taxon:4113"
/clone="STMHZ22"
/clone_lib="Generation of a set of potato cDNA clones for
microarray analyses mixed potato tissues"
/tissue_type="mixed tissues"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,

```

axillary buds of stem explants, petioles, germinating eyes ' tubers, or roots."				
BASE COUNT	287	a	112	c
ORIGIN	157	g	291	t

Query Match	6.2%	Score 125.8;	DB 14;	Length 847;
Best Local Similarity	83.9%	Pred. No. 5.1e-18;		
Matches 167; Conservative	0;	Mismatches 27;	Indels 5;	Gaps 2;

OY	31	ACGTAAGCTTTCGTCTCCTTCAACCCTTGATATATATATAATAAATTATCATTAATTAGTAGTAA	90
Db	418	AGGTTTGCTTACTTCTACCCCTTGATATATAT - - AATAATATCATTTAATTAGTAGTAA	474
OY	91	TATAATATTTCAAATATTTTTTTTCAAATAAAGAAGATGTAGTATATAGCAATTGCCTTTC	150
Db	475	AATAATATTTCCAT - TTTTTCCAAATAAAGAAGATGTAATTTAAGCAATTGCCTTTC	532
OY	151	TGTAGTTTATTAAGTGTGTATATTTTAATTATATTAACCTTTCTAATATATATGACCAAAATTTG	210
Db	533	TGTAGTTTATTAAGTGTGTATATTTTAATTATTACTTTTCTAATATATATGAGCAAAACATG	592
OY	211	TTGATGTGCAGGTACGGCG 229	
Db	593	GTCATGTTTAGGGAAGGG 611	

RESULT 4			
CNS04POX			
LOCUS	CNS04POX	621 bp	DNA
DEFINITION	Tetraodon nigrroviridis genome survey sequence T7 end of clone		
			linear
			GSS 24-MAY-2000

ACCESSION	AL301578
VERSION	AL301578.1
KEYWORDS	GSS; genome survey sequence.
SOURCE	<i>Tetraodon nigroviridis</i> .
ORGANISM	<i>Tetraodon nigroviridis</i> .

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE	AUTHORS	TITLE
1	(bases 1 to 621)	
	Roest-Crollius, H., Jalllon, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fitzames, C., Wincker, P., Brottier, P., Quetlier, F., Saurin, W. and Weissenbach, J.	
	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence	

JOURNAL REFERENCE AUTHORS	
Unpublished	
2 (bases 1 to 621)	
Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C., Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.	

TITLE	JOURNAL	REFERENCE
Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>	Unpublished	3 (bases 1 to 621)

AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000)
COMMENT	This sequence is a sing

**COMMENT**

This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

```

FEATURES      Location/Qualifiers
source        1. .621

```

```
/organism="Tetraodon nigroviridis"  
/db_xref="taxon:99883"  
/clone="127N16"  
/clone_11b="G"
```

BASE COUNT		ORIGIN	
90 a	127 c	146 g	191 t
		67 others	

Query Match	3.0%;	Score 60.4;	DB 17;	Length 621;
Best Local Similarity	41.5%;	Pred. No. 0.0036;		
Matches 151;	Conservative 30;	Mismatches 183;	Indels 0;	Gaps 0;

[illegible]

QY	416	GCCC	419
Db	417	GGCC	420

RESULT 5	CNS00DKY/c	LOCUS	DEFINITION
	CNS00DKY	928 bp	DNA linear GSS 04-JUN-1999
	Drosophila melanogaster	genome survey sequence T7 end of BAC #	
	BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		

ACCESSION AL071865  
VERSION AL071865.1. GI:4948170  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
1 (bases 1 to 928)	
Genoscope.	
Direct Submission	
Submitted (02-JUN-1999)	Genoscope - Centre National de Sequencage :

COMMENT

Determination of this BAC end sequence was carried out as part of a

- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

COMMENT

determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://bacpac.med.buffalo.edu/drosophila.bac.htm>.

```

FEATURES
  source      location/qualifiers
1. . 928
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR27A24"

```





	KEYWORDS
SOURCE ORGANISM	GSS. Drosophila melanogaster.
REFERENCE	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epiphydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 986)
AUTHORS TITLE JOURNAL COMMENT	Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr) - Web : www.genoscope.cns.fr ) determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobac11.
FEATURES	. source Location/Qualifiers 1..986 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_1lb="BACN17H15" /plasmid="pbelOBAC11" /note="end : SP6"
BASE COUNT ORIGIN	302 a 134 c 178 g 296 t 76 others
Query Match Best Local Similarity Matches	2.8%; Score 57.4; DB 17; Length 986; 45.1% ; Pred. No. 0.019; 73; Conservative 31; Mismatches 58; Indels 0; Gaps 0;
OY Db OY Db OY Db	56 ATATATATAAATTATCATTAAGTAGTAAATAAATTTCAATATTTTTTC A 115 983 MMWTATATWTTTMTTWTTTWWTTAAAWMAAATTATATATAAMWAATANA 924 116 AAATAAGAATGTAGTATATGCAATGGCTTGCTGTACTTATTAAGTGTATATTTT 175 923 AAAAAMAAMMWAATATATATATATATWTTTTTTTTTTTMMWTTTWTATTTATTT 864
OY Db	176 AATTATATACCTTTCTAATATATGCACAATTTGTTGATGT 217 863 TTTTWTWTTTWTATATMWMTATATTAAMAATTWTTT 822
RESULT LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM	9 CNS017SY 1101 bp DNA linear GSS 26-JUL-1999 CNS017SY Drosophila melanogaster genome survey sequence SP6 end of BAC AL108460 BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence. ALI08460 ALI08460.1 GI:5628764 GSS. Drosophila melanogaster. Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epiphydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101) Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr) - Web : www.genoscope.cns.fr ) determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC
REFERENCE AUTHORS TITLE JOURNAL COMMENT	d'Etupe du Polymorphisme Humain) with funding provided by a MRC

Project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

Location/Qualifiers  
1. 1101  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACN37L08"  
/clone\_1b="DrosBAC"  
/plasmid="pBelobAC11"  
/note="end : SP6"

BASE COUNT 254 a 176 c 160 g 152 t 359 others

ORIGIN

Query Match 2.88; Score 56.4; DB 17; Length 1101;  
Best Local Similarity 19.38; Pred. No. 0.033;  
Matches 121; Conservative 204; Mismatches 303; Indels 0; Gaps 0;

```

QY 58 ATATATATATATATATATATATATAGTAGTATATATATATATATTTTCAATATTTTTCAAA 117
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 459 ATACAAAAAAATTTAAAAATAARAAGTATWATWAAATWAGNATTTWATNTWNTNAANN 518

QY 118 ATAAAAGAATGTAGTATATAGCAATTGCTTTCTGTAGTTTATAAGTGTATATTTAA 177
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 519 AAANAANATNNANTTATWATWANTWNTTTWNTTNTTNTTNTTNTTNTTNTTNTTNTT 578

QY 178 TTTTAACTTTTCTTAATATATGACCAAAATTTGTTGATGTCAGGTACGGGGTGTAC 237
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 579 NANTTWTATTTTNNANAANTTTATABCTCKBKMNCGSCAGCAGCTSTKCCCACAC 638

QY 238 GACCGTCAGTCGCGGAGCGCGAGAATTCGAGCGCAGCAAGCCACGACACAGCGTAGC 297
   | : | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 639 GGSAGCGGRRKGCACAKSAGMSGCGRSGSGSSCGGKRAKGVRRGCCAGGASCA 698

QY 298 GCCAACGAAAGACAAAGCGCGCCGACCTTCAGCGCGAAGTCGAGCGGAGGGGGCGGTC 357
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 699 CMAADCGGCCAKMACCCSSSSSSSSSCASTSSSASRGVSSCACGSGSGGASACGAS 758

QY 358 AGGTCGTCGGGCATTTACGCGAAGCGCGGCGACGTCGCGGCTTGCGGACGGCGAGCGC 417
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 759 GAGGSGRGGSGCCASSGVCGAACSSASCSASMCGVSSGSSCSASCSCGCVSSCSAV 818

QY 418 CCGGAGTTCGAACGCATCTGAACGAATGCCGCGCGCGGCGCTCAACATGATCATTTTC 477
   : : | : | : | : | : | | | | | | | | | | | | | | | | | | | | | | |
Db 819 SASASSVMSKVASAVASCAVASGMSAGAVSSSCRSSVASSVSAASAASSSSSSSSSS 878

QY 478 TATGACGTGTGCGGCTTCTCGCGCTGAAGTCATGACGCGCATTCGATTGTCTCGAA 537
   : | : | : | : | : | : | | | | | | | | | | | | | | | | | | | | | |
Db 879 VVSAAAVSASSSSSASSASMAVAAAAYVSSVSVAASSSSSSSSSASVVSASVAA 938

QY 538 TTGCTCGCCCTGGGCGTGACGATGTTTCCACTCAGGAAGCGCTTCCGGCAGGGAAC 597
   : : | : | : | : | : | | | | | | | | | | | | | | | | | | | | | | |
Db 939 SASVSSSSSSSVSTSSASVSVSAVSMASVSSASSSSSSVSVVAVAAASAAAAA 998

QY 598 GTCATGACCTGATTCACCTGATTAATGCGGCTCGACGCGTCGACCAAGAATCTTGGCTG 657
   | : | : | : | | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 999 AAAAAAASSSASAVAVVSSASSSSSSSSSSSSSSSVSSSVSSSVSSSVSSVVA 1058

QY 658 AAGTCGGCGAAGATTCTCGACACGAAGA 685
   : | : | : | : | : | | | | | | | | | | | | | | | | | | | | | | |
Db 1059 VASASASASVSSVAAAAMAAAAASA 1086

```

RESULT 10  
CNS00T74/c 780 bp DNA linear GSS 28-JUN-1999  
DEFINITION Arabidopsis thaliana genome survey sequence SP6 end of BAC T3K15 of  
TAMU library from strain Columbia of Arabidopsis thaliana, genomic  
survey sequence.  
ACCESSION AL089326 GI:5290466  
VERSION AL089326.1  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana.

ORGANISM	Arabidopsis thaliana					
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.					
AUTHORS	1 (bases 1 to 780)					
JOURNAL	Salanoubat,M., Choisine,N., Artiguenave,F., Brotlier,P., Wincker,P., Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.					
REFERENCE	Unpublished					
AUTHORS	2 (bases 1 to 780)					
TITLE	Genoscope.					
JOURNAL	Direct Submission					
AUTHORS	Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr					
FEATURES	- Web : www.genoscope.cns.fr)					
source	Location/Qualifiers					
	1..780					
	/organism="Arabidopsis thaliana"					
	/strain="Columbia"					
	/db_xref="taxon:3702"					
	/clone="T3K15"					
	/clone_1lb="TAMU"					
	/note="end : SP6"					
BASE COUNT	282 a 143 c 141 g 202 t 12 others					
ORIGIN						
Query Match	2.8%; Score 56; DB 17; Length 780;					
Best Local Similarity	61.8%; Pred. No. 0.038;					
Matches	89; Conservative 0; Mismatches 55; Indels 0; Gaps 0;					
OY	39 TTCTGCTTCTACCTTGATATATATATAATTATTCATTAAATTAGTAGTAATATATAT 98					
Db	312 TTTTCTTTCTTTATATATATATATAAATAATAATAATAATAATAATAATAAAAAATTAC 253					
OY	99 TTCAAATATTTTTTTCAAATAAAGAATGTAGTATATAGCAATTCCTTTCTGTAGTTT 158					
Db	252 TATATATTTTTTTATTAGAAAAAATAAGCTGGAATTTGCTACCTTTTATTATTTTAT 193					
OY	159 ATAAGTGTGTATATTTTAATTTAT 182					
Db	192 TGAAATTTGTATTTTTTTTTTTTT 169					
RESULT 11						
CNS01038						
LOCUS	CNS01038 1201 bp DNA linear GSS 26-JUL-1999					
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC					
VERSION	BACN03017 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.					
KEYWORDS	AL098462					
SOURCE	AL098462.1 GI:5610073					
ORGANISM	GSS.					
	Drosophila melanogaster.					
	Drosophila melanogaster					
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.					
REFERENCE	1 (bases 1 to 1201)					
AUTHORS	Genoscope.					
TITLE	Direct Submission					
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr					
COMMENT	- Web : www.genoscope.cns.fr)					
	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.					
FEATURES						
source	Location/Qualifiers					
	1..1201					
	/organism="Drosophila melanogaster"					

BASE COUNT	187	a	158	c	150	g	394	t	312	others
ORIGIN										
Query Match	2.8%; Score 56; DB 17; Length 1201;									
Best Local Similarity	37.2%; Pred. No. 0.041;									
Matches	74;	Conservative	49;	Mismatches	76;	Indels	0;	Gaps	0	
QY	8	AAGGGGTGTGACCGGGGTGATACGTAAGTTCTGCTTCTACCTTGATATATATATAA	67							
Db	940	ATGRGKGTKTDKKDDGTTTKGATTTKGTGDGATWAAAMDDKDKKKKAWRKDAKWRWAA	999							
QY	68	TAATTATCATTAATAGTAGTAATATATATATTTTCAATATTTTTCAAAATAAAGAAT	127							
Db	1000	KGWAAMWMAANAGAAAAATWAWTATAAATDRDMATTTTTTTTTATTTTWTWTTAAAW	1059							
QY	128	GTAGTATATAGCAATGCTTTTCTGTAGTTTATTAAGTGTGTATATTTTAATTTAATTA	187							
Db	1060	ATWWATATADAAAWATATWKAAMTWTTAAATANTWTTTWTWTWTTWTTWAAATTTT	1119							
QY	188	TTCTAATATATGACCAAAA	206							
Db	1120	TTTTTAAATATATAAAAAA	1138							
RESULT 12										
CNS00ZAL										
LOCUS	CNS00ZAL	1101 bp	DNA	linear	GSS	26-JUL-1999				
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN01120 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.									
ACCESSION	AL097431									
VERSION	AL097431.1	GI:5609042								
KEYWORDS	GSS.									
SOURCE	Drosophila melanogaster.									
ORGANISM	Drosophila melanogaster									
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.									
AUTHORS	1 (bases 1 to 1101)									
TITLE	Genoscope.									
JOURNAL	Direct Submission									
COMMENT	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)									
FEATURES	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.									
Source	Location/Qualifiers									
	1..1101									
	/organism="Drosophila melanogaster"									
	/db_xref="taxon:7227"									
	/clone="BACN01120"									
	/clone_lib="DrosBAC"									
	/plasmid="pBelOBAC11"									
	/note="end : SP6"									
BASE COUNT	391	a	216	c	197	g	149	t	148	others
ORIGIN										
Query Match	2.7%; Score 55.8; DB 17; Length 1101;									
Best Local Similarity	44.3%; Pred. No. 0.045;									
Matches	74;	Conservative	32;	Mismatches	61;	Indels	0;	Gaps	0;	
QY	29	ATACGTAAGTTCTGCTTACCTTTGATATATATATATAATTAATTAATTAATTAAGT	88							

[illegible]

Db	704	TATTWWWTATWTTATATATAATTTTWTATTATATATWATWATATTATTTAT	657
RESULT 14	CNS003BD/C	LOCUS	1101 bp DNA linear GSS 03-JUN-1999
DEFINITION	CNS003BD		Drosophila melanogaster genome survey sequence TET3 end of BAC #
ACCESSION	BACR08K08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
VERSION	AL064091		
KEYWORDS	AL064091.1 GI:4941847		
SOURCE	GSS.		
ORGANISM	Drosophila melanogaster.		
REFERENCE	Drosophila melanogaster.		
AUTHORS	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
TITLE	1 (bases 1 to 1101)		
JOURNAL	Genoscope.		
COMMENT	Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library">http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library</a> was prepared by Kazutoyo Osogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .		
FEATURES	location/Qualifiers		
source	1. .1101		
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
	/clone="BACR08K08"		
	/clone_lib="RPCI-98"		
	/note="end : TET3"		
BASE COUNT	395 a 120 c 103 g 334 t 149 others		
ORIGIN			
Query Match	2.7%; Score 55.4; DB 17; Length 1101;		
Best Local Similarity	50.3%; Pred. No. 0.056;		
Matches	74; Conservative 21; Mismatches 52; Indels 0; Gaps 0;		
QY	52 TTTCATATATATATAATATATCATTAATTAGTAGTAATATAATATTTCAATATTTT 111		
Db	1006 TTTTATTTTTTTTTTWWTATTTATAATWATWATATAAAATATTWTTWTTAAATWTTWTT 947		
QY	112 TTCAAAATAAAGAATGTAGTATATAGCAATTGCTTTTCTGTAGTTTATAAGTGTATA 171		
Db	946 ARKAAAWAAMWATWTTWATTTTATTTTAAATTTTTTTTWTATWTTTAAWTTTAAWTTAAAW 887		
QY	172 TTTTAAATTATAACTTTTCTAATATAT 198		
Db	886 WTATWATATWATTTATTTTATTTATTTT 860		
RESULT 15	BH179465		
LOCUS	BH179465	975 bp	DNA linear GSS 29-OCT-2001
DEFINITION	014_P10-21 SMBAC1 Schistosoma mansoni genomic clone 014P10 5', DNA sequence.		
ACCESSION	BH179465		
VERSION	BH179465.1 GI:16280351		



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 17:41:53 ; Search time 107 Seconds  
(Without alignments)  
5821.126 Million cell updates/sec

Title: US-09-940-550A-9

Perfect score: 2031

Sequence: 1 atgacacacaggggtgtgtac.....cggaagacgtagcggttag 2031

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1812.6	89.2	2091	6 5190871-3	Patent No. 5190871
2	1809.4	89.1	3401	6 5190871-1	Patent No. 5190871
3	195	9.6	2633	1 US-08-452-267-2	Sequence 2, Appli
4	195	9.6	2633	3 US-09-123-644-2	Sequence 2, Appli
5	195	9.6	5534	1 US-08-452-267-3	Sequence 3, Appli
6	195	9.6	5534	3 US-09-123-644-3	Sequence 3, Appli
7	194	9.6	1701	4 US-09-080-625-1	Sequence 1, Appli
8	194	9.6	1701	4 US-09-695-782-1	Sequence 1, Appli
9	194	9.6	3336	4 US-09-080-625-2	Sequence 2, Appli
10	194	9.6	3336	4 US-09-695-782-2	Sequence 2, Appli
11	194	9.6	3877	4 US-09-080-625-4	Sequence 4, Appli
12	194	9.6	3877	4 US-09-695-782-4	Sequence 4, Appli
13	190.2	9.4	5560	3 US-08-817-188-5	Sequence 5, Appli
14	190	9.4	1034	4 US-09-367-293-2	Sequence 2, Appli
15	52.2	2.6	1926	4 US-09-249-585A-4	Sequence 4, Appli
16	52.2	2.6	1931	2 US-09-130-114-2	Sequence 2, Appli
17	51.2	2.5	19124	2 US-08-487-826B-13	Sequence 13, Appli
18	46.8	2.3	19124	2 US-08-487-826B-13	Sequence 13, Appli
19	46.6	2.3	11485	4 US-09-410-464-9	Sequence 9, Appli
20	45	2.2	722	4 US-08-998-416-780	Sequence 780, App
21	45	2.2	727	4 US-08-998-416-1011	Sequence 1011, App
22	44.8	2.2	8700	2 US-08-392-625-16	Sequence 16, Appli
23	44.8	2.2	8700	2 US-08-466-961A-16	Sequence 16, Appli
24	44.8	2.2	8700	2 US-08-645-193B-18	Sequence 18, Appli
25	44.4	2.2	5923	4 US-09-064-922-3	Sequence 3, Appli
26	44	2.1	2960	3 US-08-913-842-3	Sequence 3, Appli
27	43.4	2.1	1850	3 US-08-617-860B-32	Sequence 32, Appli

28	43.4	2.1	4098	2 US-08-605-106-4	Sequence 4, Appli
29	43	2.1	4257	2 US-08-690-473-1	Sequence 1, Appli
30	43	2.1	4257	4 US-09-259-821A-1	Sequence 1, Appli
31	43	2.1	4257	4 US-08-843-659-1	Sequence 1, Appli
32	43	2.1	5852	1 US-07-867-106-2	Sequence 2, Appli
33	43	2.1	12001	1 US-08-458-568A-11	Sequence 11, Appli
34	42.6	2.1	6265	4 US-09-129-112-3	Sequence 3, Appli
35	42.2	2.1	1431	4 US-09-316-083-2	Sequence 2, Appli
36	42	2.1	615	4 US-08-998-416-186	Sequence 186, App
37	41.6	2.0	7218	1 US-08-232-463-14	Sequence 14, Appli
38	41.4	2.0	7815	4 US-09-102-528-28	Sequence 28, Appli
39	41.4	2.0	20235	1 US-07-642-734C-3	Sequence 3, Appli
40	41.4	2.0	20235	3 US-08-439-009A-3	Sequence 3, Appli
41	41.4	2.0	168575	4 US-09-426-290-1	Sequence 1, Appli
42	41	2.0	2960	3 US-08-913-842-3	Sequence 3, Appli
43	40.8	2.0	3618	1 US-07-872-678A-36	Sequence 36, Appli
44	40.8	2.0	4403765	4 US-09-103-840A-2	Sequence 2, Appli
45	40.8	2.0	4411529	4 US-09-103-840A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
5190871-3  
Patent No. 5190871  
APPLICANT: COX, KAREN L.; KUHSTOSS, STUART A.; RAO, R. NAGARAJA  
; RICHARDSON, MARK A.; SCHONER, BRIGITTE E.; SENO, EUGENE T.  
; TITLE OF INVENTION: USE OF THE SITE-SPECIFIC INTEGRATING  
; FUNCTION OF PHAGE C31  
; NUMBER OF SEQUENCES: 3  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/364,959  
; FILING DATE: 12-JUN-1989  
; SEQ ID NO:3:  
; LENGTH: 2091  
5190871-3

Query Match 89.2% Score 1812.6; DB 6; Length 2091;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1815; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	213	GATGTGACGATGCGGGGTGCTTACGACCGTCACTCGCGGCGGAGAGATTGAGCGC	272
DB	254	GGTGACACGATGCGGGGTGCTTACGACCGTCACTCGCGGCGGAGAGATTGAGCGC	313
QY	273	AGCAAGCCGACGACACAGCGTAGCGCCCAAGAGAGAGCGCGGAGAGATTGAGCGC	332
DB	314	AGCAAGCCGACGACACAGCGTAGCGCCCAAGAGAGAGCGCGGAGAGATTGAGCGC	373
QY	333	AGTCGAGCGGCGGCGGGGCGGTTGAGGTTGCGGGGATTCAGCGGCGGCGC	392
DB	374	AGTCGAGCGGCGGCGGGGCGGTTGAGGTTGCGGGGATTCAGCGGCGGCGC	433
QY	393	GTCGCGTTCGCGGAGCGGCGGAGCGCGGAGTTGAGAGCGATCTGAAGAGATGCGCGC	452
DB	434	GTCGCGTTCGCGGAGCGGCGGAGCGCGGAGTTGAGAGCGATCTGAAGAGATGCGCGC	493
QY	453	CGGCGGCTCAACATGATGATGTCATGAGCTGTGCGGCTTCGCGGCTGAAGATCAT	512
DB	494	CGGCGGCTCAACATGATGATGTCATGAGCTGTGCGGCTTCGCGGCTGAAGATCAT	553
QY	513	GGACGCGATTCGATGTCCTCGGAATGCTCGCCCTGGGCGGTGACGATGTTTCCACTCA	572
DB	554	GGACGCGATTCGATGTCCTCGGAATGCTCGCCCTGGGCGGTGACGATGTTTCCACTCA	613
QY	573	GGAAGCGCTTCGCGGAGGAGGAAACGTCATGACCTGATTCACCTGATTCGCGGCTGCA	632
DB	614	GGAAGCGCTTCGCGGAGGAGGAAACGTCATGACCTGATTCACCTGATTCGCGGCTGCA	673
QY	633	CGCGTTCGACAAAGATCTTCGCTGAAGTGGCGGAGATTCGACAGAGAACCTTCA	692
DB	674	CGCGTTCGACAAAGATCTTCGCTGAAGTGGCGGAGATTCGACAGAGAACCTTCA	733



QY 693 GCGCAATTGGGGGCTACGTCGGCGGGAAGGCGCCTTACGGCTTCGAGCTTGTTCGA 752  
DB 734 GCGCAATTGGGGGCTACGTCGGCGGGAAGGCGCCTTACGGCTTCGAGCTTGTTCGA 793  
QY 753 GACGAAGAGATCACGCGCAACGCGCGGAATGTCATATGTCATCAACAAGCTTGGCGA 812  
DB 794 GACGAAGAGATCACGCGCAACGCGCGGAATGTCATATGTCATCAACAAGCTTGGCGA 853  
QY 813 CTCGACCACTCCCTTACCGGACCCCTTCGAGTTCGAGCCCGACGTAATCCGGTGTGTG 872  
DB 854 CTCGACCACTCCCTTACCGGACCCCTTCGAGTTCGAGCCCGACGTAATCCGGTGTGTG 913  
QY 873 GCGTAGATCAAGACGCAACAACACCTTCCCTTCAAGCCGGGACGTCAAGCCGCCATCA 932  
DB 914 GCGTAGATCAAGACGCAACAACACCTTCCCTTCAAGCCGGGACGTCAAGCCGCCATCA 973  
QY 933 CCGGGGACGATCAGCGGGCTTGTGAAGCGATGACGCTGACGCGGTGCCGACCGGGG 992  
DB 974 CCGGGGACGATCAGCGGGCTTGTGAAGCGATGACGCTGACGCGGTGCCGACCGGGG 1033  
QY 993 CGAGACGATTGGGAGAAGAACCGCTTCAAGCCCTGGAGCCCGCAACCGTTATGCGAAT 1052  
DB 1034 CGAGACGATTGGGAGAAGAACCGCTTCAAGCCCTGGAGCCCGCAACCGTTATGCGAAT 1093  
QY 1053 CCTTCGGGACCCGCTATTGCGGGCTTCGCCCTGAGTGATCTACAAGAAGCCGGA 1112  
DB 1094 CCTTCGGGACCCGCTATTGCGGGCTTCGCCCTGAGTGATCTACAAGAAGCCGGA 1153  
QY 1113 CGGCACGCCGACCAAGATTTAGGGTTACCGCATTCAGCGCGACCCGATCAGCTCCG 1172  
DB 1154 CGGCACGCCGACCAAGATTTAGGGTTACCGCATTCAGCGCGACCCGATCAGCTCCG 1213  
QY 1173 GCCGGTCGAGCTTGAATGCGGACCGATCATCGAGCCCGTGAATGATGAGCTTCAGGC 1232  
DB 1214 GCCGGTCGAGCTTGAATGCGGACCGATCATCGAGCCCGTGAATGATGAGCTTCAGGC 1273  
QY 1233 GTGGTTGACGCGGAGGGGCGCGGAGGGGCTTCCCGGGGCAAGCCATTCTGCCG 1292  
DB 1274 GTGGTTGACGCGGAGGGGCGCGGAGGGGCTTCCCGGGGCAAGCCATTCTGCCG 1333  
QY 1293 CATGACAAGCTGTACTGCGAGTGTGGCGCCGCTCATGATTCGAAGCGCGGGGAAGATC 1352  
DB 1334 CATGACAAGCTGTACTGCGAGTGTGGCGCCGCTCATGATTCGAAGCGCGGGGAAGATC 1393  
QY 1353 GATCAAGACTTCTTACCGCTGCCCTGCCGGAAGGTGCTGACCCGCTCCGACCTGGGA 1412  
DB 1394 GATCAAGACTTCTTACCGCTGCCCTGCCGGAAGGTGCTGACCCGCTCCGACCTGGGA 1453  
QY 1413 GCACGAAGGACGTGCAACGTGACATGGCGGACCTGCAAGTTCGTTGGGGAAGCAT 1472  
DB 1454 GCACGAAGGACGTGCAACGTGACATGGCGGACCTGCAAGTTCGTTGGGGAAGCAT 1513  
QY 1473 CTTCACAAGATCAGGACCGCGGAAGGCGAGGAAGAGCTTGGCGCTTGTGGGAAGC 1532  
DB 1514 CTTCACAAGATCAGGACCGCGGAAGGCGAGGAAGAGCTTGGCGCTTGTGGGAAGC 1573  
QY 1533 CGCCCGACGCTTCGGCAAGCTCAGAGGCGCTGAGAAAGCGGCGAAGCGGCAACCT 1592  
DB 1574 CGCCCGACGCTTCGGCAAGCTCAGAGGCGCTGAGAAAGCGGCGAAGCGGCAACCT 1633  
QY 1593 TGTTCGGGAGCGCGCGGCGCTGAAAGCGCTTGAAGAGCTGTAGAAAGCGCGCGG 1652  
DB 1634 TGTTCGGGAGCGCGCGGCGCTTGAAGAGCTGTAGAAAGCGCGCGG 1693  
QY 1653 AGCGCGGTACGACGACCGGCTTGGCAGGAAGCACTTCGGAAGCAAGCAGCGGTGAC 1712  
DB 1694 AGCGCGGTACGACGACCGGCTTGGCAGGAAGCACTTCGGAAGCAAGCAGCGGTGAC 1753  
QY 1713 GCTCCGGCAGCAAGGGGGAAGAGCGGCTTGGCAACTTGAAGCGCGCGAAGCCCGAA 1772  
DB 1754 GCTCCGGCAGCAAGGGGGAAGAGCGGCTTGGCAACTTGAAGCGCGCGAAGCCCGAA 1813

QY 1773 GCTTCCCTTGACCAATGTTCCCGGAGACGCGGACGCTGACCCGACCGCCCTAAGTC 1832  
DB 1814 GCTTCCCTTGACCAATGTTCCCGGAGACGCGGACGCTGACCCGACCGCCCTAAGTC 1873  
QY 1833 GTGTGGGGCGCGCTCAGTAGACGCAAGCGGCTGTTGCTGGGCTCTTCGTAGCAA 1892  
DB 1874 GTGTGGGGCGCGCTCAGTAGACGCAAGCGGCTGTTGCTGGGCTCTTCGTAGCAA 1933  
QY 1893 GATCGTTGTCAGGAAGTCTGACTACGGGCAAGGGGCGAGGGAAGCCCATCGAAGCGCGC 1952  
DB 1934 GATCGTTGTCAGGAAGTCTGACTACGGGCAAGGGGCGAGGGAAGCCCATCGAAGCGCGC 1993  
QY 1953 TTCGATCACGTGGGCGGAAGCCCGGACCGACGACGACGAGCAAGAGCCGACGCGCAC 2012  
DB 1994 TTCGATCACGTGGGCGGAAGCCCGGACCGACGACGACGAGCAAGAGCCGACGCGCAC 2053  
QY 2013 GGAAGACGTAGCGCGTAG 2031  
DB 2054 GGAAGACGTAGCGCGTAG 2072

RESULT 2  
5190871-1  
; Patent No. 5190871  
; APPLICANT: COX, KAREN L.; KUHSTOSS, STUART A.; RAO, R. NAGARAJA  
; RICHARDSON, MARK A.; SCHONER, BRIGITTE E.; SENO, EUGENE T.  
; TITLE OF INVENTION: USE OF THE SITE-SPECIFIC INTEGRATING  
; FUNCTION OF PHAGE C31  
; NUMBER OF SEQUENCES: 3  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/364, 959  
; FILING DATE: 12-JUN-1989  
; SEQ ID NO:1:  
; LENGTH: 3401  
5190871-1

Query Match 89.1%; Score 1809.4; DB 6; Length 3401;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1813; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 213 GATGTCAAGGTACGCGGGTCTTACGACCGTCACTGCGGCGAGCGCGAATTCGAGCGC 272  
DB 532 GGTGACACGTACGCGGGTCTTACGACCGTCACTGCGGCGAGCGCGAATTCGAGCGC 591  
QY 273 AGCAAGCCAGCGACACAGCTAGCGGCAAGCAAGAGCGGCGGCTTCAGCGCGA 332  
DB 592 AGCAAGCCAGCGACACAGCTAGCGGCAAGCAAGAGCGGCGGCTTCAGCGCGA 651  
QY 333 AGTCGAGCGGACGCGGCGGCTTCAAGTTCGCGGCTTCAGCGGCAAGCGGCGGAC 392  
DB 652 AGTCGAGCGGACGCGGCGGCTTCAAGTTCGCGGCTTCAGCGGCAAGCGGCGGAC 711  
QY 393 GTCGGCTTCGGGACGCGGAGCGCGCGGCTTCGAAGCATCTGAACGAATGCGCGC 452  
DB 712 GTCGGCTTCGGGACGCGGAGCGCGCGGCTTCGAAGCATCTGAACGAATGCGCGC 771  
QY 453 CGGGCGCTCAACATGATCATTTGTATGACGTGTGCGGCTTTCGCGGCTGAAGGTAT 512  
DB 772 CGGGCGCTCAACATGATCATTTGTATGACGTGTGCGGCTTTCGCGGCTGAAGGTAT 831  
QY 513 GGAAGCATTCGATTTGCTCGGAATTCGCGGCTGGGCGGTGACGATTTTCCACTCA 572  
DB 832 GGAAGCATTCGATTTGCTCGGAATTCGCGGCTGGGCGGTGACGATTTTCCACTCA 891  
QY 573 GGAAGCATTCGCGGCAAGGAACGTCATGACCTGATTCACCTGATTAAGCGGCTGA 632  
DB 892 GGAAGCATTCGCGGCAAGGAACGTCATGACCTGATTCACCTGATTAAGCGGCTGA 951  
QY 633 CGGCTGCAACAAGATCTTCGCTGAGTGGCGAAGATTCGACAGCAAGAACCTTCA 692  
DB 952 CGGCTGCAACAAGATCTTCGCTGAGTGGCGAAGATTCGACAGCAAGAACCTTCA 1011  
QY 693 GCGCAATTGGGCGGCTACGTGCGGCGGGAAGGCGCCTTACGGCTTCGAGCTTGTTCGA 752



QY	90	ATATAATATTCCAATATTTTTCAAAATAAAGAATGTAGTATATAGCAATTTGCTTTT	149
Db	753	ATATAATATTCCAATATTTTTCAAAATAAAGAATGTAGTATATAGCAATTTGCTTTT	812
QY	150	CTGTAGTTTATAAGTGTATATTTTAATTTAACCTTTTCTAATATATGACCAAAATTT	209
Db	813	CTGTAGTTTATAAGTGTATATTTTAATTTAACCTTTTCTAATATATGACCAAAATTT	872
QY	210	GTGTATGTGCAGGTA 224	
Db	873	GTGTATGTGCAGGTA 887	

RESULT 4  
US-09-123-644-2  
; Sequence 2, Application US/09123644  
; Patent No. 6127606

```

: GENERAL INFORMATION:
: APPLICANT: Bennett, Malcolm
: APPLICANT: May, Sean
: APPLICANT: Ramsay, Nicola
: TITLE OF INVENTION: Method of Using Transactivation Proteins to
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Barnes & Thornburg
: STREET: 11 South Meridian
: CITY: Indianapolis
: STATE: IN
: COUNTRY: USA
: ZIP: 46204
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/123,644
: FILING DATE: 28-JUL-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Breen, John P.
: REGISTRATION NUMBER: 38,833
: REFERENCE/DOCKET NUMBER: 6653-60788
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (317) 231-7745
: TELEFAX: (317) 231-7433
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2633 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: IMMEDIATE SOURCE:
: CLONE: pUMIGIT(insert)
:
US-09-123-644-2

```

Query Match	9.6%;	Score 195;	DB 3;	Length 2633;
Best Local Similarity	100.0%;	Pred. No. 3.6e-38;		
Matches 195; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0.0;

QY	30	TACGTAAGTTCTGCCTTCTACCCTTGATATATATATAAATATTCATTAAATTAGTAGTA	89
Db	693	TACGTAAGTTCTGCCTTCTACCCTTGATATATATATAAATATTCATTAAATTAGTAGTA	752
QY	90	ATATAATATTCCAATATTTTTTTCCAAAATAAAGAATGTAGTATATAGCAATTGCTTTT	149
Db	753	ATATAATATTCCAATATTTTTTTCCAAAATAAAGAATGTAGTATATAGCAATTGCTTTT	812
QY	150	CTGTAGTTTATAGGTGTATATTTTAAATTATTAACCTTTGCCAATATATGACCAAATTT	209

Db	813	CTGTAGTTTATTAAGTGTGTATATTTTAAATTATTAACCTTTCTAATATATGACCAAAATTT	872
QY	210	GTGTATGTCAGGTA	224
Db	873	GTGTATGTCAGGTA	887

RESULT 5  
US-08-452-267-3  
; Sequence 3, Application US/08452267  
; Patent No. 5801027

```

GENERAL INFORMATION:
APPLICANT: Bennett, Malcolm
APPLICANT: May, Sean
APPLICANT: Ramsay, Nichola
TITLE OF INVENTION: Control of Genes In Transgenic Plants
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Barnes & Thornburg
STREET: 11 South Meridian
CITY: Indianapolis
STATE: IN
COUNTRY: USA
ZIP: 46204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,267
FILING DATE: 26-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Breen, John P.
REGISTRATION NUMBER: 38,833
REFERENCE/DOCKET NUMBER: 6653-25744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 231-7745
TELEFAX: (317) 231-7433
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5534 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: pUMIGIT

```

Query Match	9.6%;	Score 195;	DB 1;	Length 5534;
Best Local Similarity	100.0%;	Pred. No. 5.1e-38;		
Matches 195; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	30	TACGTAAGTTTCTGCTTCTACCTTTGATATATATATAATAATATCATTAATTAGTAGTA	89
- Db	688	TACGTAAGTTTCTGCTTCTACCTTTGATATATATATAATAATATCATTAATTAGTAGTA	747
QY	90	ATATAATATTCCAATATTTTTTCCAATAAAGAATGTAGTATATAGCAATTGCTTTT	149
Db	748	ATATAATATTCCAATATTTTTTCCAATAAAGAATGTAGTATATAGCAATTGCTTTT	807
QY	150	CTGTAGTTTATAAGTGTGTATATTTTAATTATTAACCTTTTCTAATATATGACCAAAATTT	209
Db	808	CTGTAGTTTATAAGTGTGTATATTTTAATTATTAACCTTTTCTAATATATGACCAAAATTT	867
QY	210	GTGTATGTGCAGGTA 224	
Db	868	GTGTATGTGCAGGTA 882	

```

RESULT 6
US-09-123-644-3
; Sequence 3, Application US/09123644
; Patent No. 6127606
; GENERAL INFORMATION:
; APPLICANT: Bennett, Malcolm
; APPLICANT: May, Sean
; APPLICANT: Ramsay, Nicola
; TITLE OF INVENTION: Method of Using Transactivation Proteins to
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 South Meridian
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/123,644
; FILING DATE: 28-JUL-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Breen, John P.
; REGISTRATION NUMBER: 38,833
; REFERENCE/DOCKET NUMBER: 6653-60788
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7745
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5534 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: pUMIGIT
;
US-09-123-644-3

Query Match          9.6%; Score 195; DB 3; Length 5534;
Best Local Similarity 100.0%; Pred. No. 5.1e-38;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY      30 TACGTAAGTTTCTGCTTCTACCTTGATATATATATAAATTAATTCATTAGTAGTA 89
        |||||||
DB       688 TAGCTAAGTTTCTGCTTCTACCTTGATATATATATAAATTAATTCATTAGTAGTA 747
QY      90 ATATAATATTTCAAAATATTTTTTCAAAATAAAAGAATGTAGTATATGCAATTGCTTTT 149
        |||||||
DB       748 ATATAATATTTCAAAATATTTTTTCAAAATAAAAGAATGTAGTATATGCAATTGCTTTT 807
QY     150 CTGTAGTTTATAAGTGTGTATATTTTAATTATAACTTTTCTAATATATGACCAGCAAAATTT 209
        |||||||
DB       808 CTGTAGTTTATAAGTGTGTATATTTTAATTATAACTTTTCTAATATATGACCAGCAAAATTT 867
QY     210 GTTGATGTGCAGGTA 224
        |||||||
DB       868 GTTGATGTGCAGGTA 882

RESULT 7
US-09-080-625-1
; Sequence 1, Application US/09080625
; Patent No. 6307123
; GENERAL INFORMATION:
```

```

; APPLICANT: Kriz, Alan L.
; APPLICANT: Spencer, T. Michael
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE
; TITLE OF INVENTION: IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,625
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, Robert E.
; REGISTRATION NUMBER: P-42,628
; REFERENCE/DOCKET NUMBER: DEKM:161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1701 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-09-080-625-1
;
Query Match          9.6%; Score 194; DB 4; Length 1701;
Best Local Similarity 97.5%; Pred. No. 5.2e-38;
Matches 197; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      23 GGGTGATAGTAAGTTTCGCTTCTACCTTGATATATAAATAATTATCATTAATT 82
       11 11111111111111111111111111111111111111111111111111111
Db      370 GAGCTGAAGGTAAGTTTCGCTTCTACCTTGATATATAAATAATTATCATTAATT 429
QY      83 AGTAGTAATATAATATTTCAAAATATTTTTTCAAAATAAAGAATGTAGTATAGCAAT 142
       11111111111111111111111111111111111111111111111111111
Db      430 AGTAGTAATATAATATTTCAAAATATTTTTTCAAAATAAAGAATGTAGTATAGCAAT 489
QY      143 TGCTTTCTGTAGTTTATATAGTGATATATTTTAAATTTTAACTTTCTAATATAGACC 202
       11111111111111111111111111111111111111111111111111111
Db      490 TGCTTTCTGTAGTTTATATAGTGATATATTTTAAATTTTAACTTTCTAATATAGACC 549
QY      203 AAAATTGTGATGTGCAGGTA 224
       11111111111111111111111111111111111111111111111111111
Db      550 AAAATTGTGATGTGCAGGTA 571
;
RESULT 8
US-09-695-782-1
; Sequence 1, Application US/09695782
; Patent No. 6433252
; GENERAL INFORMATION:
; APPLICANT: KRIZ, ALAN L.
; APPLICANT: GRIEFOR, MATTHEW
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE IDENTIFICATION
; FILE REFERENCE: DEKM:161USC1
; CURRENT APPLICATION NUMBER: US/09/695,782
; CURRENT FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: 09/080,625
; PRIOR FILING DATE: 1998-05-18
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1

```

NUMBER OF SEQUENCES: 19



CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Arnold White & Durkee  
 STREET: P.O. Box 4433  
 CITY: Houston  
 STATE: TX  
 COUNTRY: USA  
 ZIP: 77210-4433  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/080,625  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Hanson, Robert E.  
 REGISTRATION NUMBER: P-42,628  
 REFERENCE/DOCKET NUMBER: DEKM:161  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (512) 418-3000  
 TELEFAX: (512) 474-7577  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3877 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear

Query Match	9.6%	Score 194;	DB 4;	Length 3877;
Best Local Similarity	97.5%	Pred. No. 7.6e-38;		
Matches 197; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0

OY	23	GGGTGATACGTAAGTTCTTGCTTCTACCCTTGGATAATATAATAATTATCAITAAATT	82
Db	1982	GAGCTGAAGAAGTAGTTCTGCTTCCTACCCTTGATATATATAAATAATTATCATTAATT	2041
OY	83	AGTAGTAATATAATATTCCAATATTTTTTTCAAATAAAGAAGTAGTATATAGCAAT	142
Db	2042	AGTAGTAATATAATATTCCAATATTTTTTTCAAATAAAGAAGTAGTATATAGCAAT	2101
OY	143	TGCTTTTCTGTAGTTATAAGTGTGTATATTTTAATTTATACTTTTCTAATATATGACC	202
Db	2102	TGCTTTTCTGTAGTTATAAGTGTGTATATTTTAATTTATAACTTTTCTAATATATGACC	2161
OY	203	AAAATTTGTTGATGTGCAGGTA	224
Db	2162	AAAAATTTGTTGATGTGCAGGTA	2183

```

RESULT 12
US-09-695-782-4
; Sequence 4, Application US/09695782
; Patent No. 6433252
; GENERAL INFORMATION:
; APPLICANT: KRIZ, ALAN L.
; APPLICANT: GRIFFOR, MATTHEW
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE IDENTIFICATION
; FILE REFERENCE: DEXM:161USC1
; CURRENT APPLICATION NUMBER: US/09/695,782
; CURRENT FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: 09/080,625
; PRIOR FILING DATE: 1998-05-18
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 4
; LENGTH: 3877
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic

```

```

; OTHER INFORMATION: Primer
US-09-695-782-4

Query Match          9.6%; Score 194; DB 4; Length 3877;
Best Local Similarity 97.5%; Pred. No. 7.6e-38;
Matches 197; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      23 GGGTGATACGTAAGTTTCGCTTCTACCTTTGATATATATAATAATTATCATTAATT 82
        |||||
Db       1982 GAGCTGAAGGTAAGTTTCGCTTCTACCTTTGATATATATAATAATTATCATTAATT 2041

QY      83 AGTAGTAATATAATATTTCAAATATTTTTCAAAATAAAGCAAGTAGTATATAGCAAT 142
        |||||
Db       2042 AGTAGTAATATAATATTTCAAATATTTTTCAAAATAAAGCAAGTAGTATATAGCAAT 2101

QY      143 TGCTTTTCTGTAGTTTATTAAGTGTGTATATTTTAATTATTAACCTTCTAATATATGACC 202
        |||||
Db       2102 TGCTTTTCTGTAGTTTATTAAGTGTGTATATTTTAATTATTAACCTTCTAATATATGACC 2161

QY      203 AAAATTTGTTGATGTGCAGGTA 224
        |||||
Db       2162 AAAATTTGTTGATGTGCAGGTA 2183

```

```

RESULT 13
US-08-817-188-5/c
Sequence 5, Application US/08817188
Patent No. 6074876
GENERAL INFORMATION:
APPLICANT: DE BLOCK, MARC
TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
FILE REFERENCE: 2121-0127P
CURRENT APPLICATION NUMBER: US/08/817,188
CURRENT FILING DATE: 1997-05-15
EARLIER APPLICATION NUMBER: PCT/EP96/03366
EARLIER FILING DATE: 1996-07-31
EARLIER APPLICATION NUMBER: EP.95401844.6
EARLIER FILING DATE: 1995-08-04
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 5560
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
OTHER INFORMATION: plasmid pTHW142
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(25)
OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
OTHER INFORMATION: pTI186S3
FEATURE:
NAME/KEY: misc_feature
LOCATION: (84)..(296)
OTHER INFORMATION: 3' g7: 3' untranslated region containing the
OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
OTHER INFORMATION: T-DNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: (318)..(869)
OTHER INFORMATION: bar: region coding for phosphinotricin
OTHER INFORMATION: acetyltransferase
FEATURE:
NAME/KEY: misc_feature
LOCATION: (830)..(2760)
OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit
OTHER INFORMATION: gene of Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_feature
LOCATION: (2765)..(3058)
OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
OTHER INFORMATION: containing polyadenylation signals

```

```

1 FEATURE:
2 NAME/KEY: misc_feature
3 LOCATION: (3059)..(5056)
4 OTHER INFORMATION: uida: region coding for beta-glucuronidase
5 FEATURE:
6 NAME/KEY: misc_feature
7 LOCATION: (4483)..(4671)
8 OTHER INFORMATION: IV2: region corresponding to the second intron of
9 OTHER INFORMATION: the ST-L51 gene
10 FEATURE:
11 NAME/KEY: misc_feature
12 LOCATION: (5067)..(5502)
13 OTHER INFORMATION: P35S: 35S promoter region of CamV
14 FEATURE:
15 NAME/KEY: misc_feature
16 LOCATION: (5533)..(5560)
17 OTHER INFORMATION: LB: left border sequence of octopline TL-DNA from
18 OTHER INFORMATION: PTIB6S3
19 FEATURE:
20 NAME/KEY: misc_feature
21 LOCATION: (5058)..(5059)
22 OTHER INFORMATION: region with unknown sequence (may contain up to 20
23 OTHER INFORMATION: nucleotides)
24 FEATURE:
25 NAME/KEY: misc_feature
26 LOCATION: (5077)..(5078)
27 OTHER INFORMATION: region with unknown sequence (may contain up to 20
28 OTHER INFORMATION: nucleotides)
29 FEATURE:
30 NAME/KEY: misc_feature
31 LOCATION: (5476)..(5479)
32 OTHER INFORMATION: region with unknown sequence (may contain up to 20
33 OTHER INFORMATION: nucleotides)
34
35 US-08-187-188-5

```

Query Match	Best Local Similarity	9.48;	Score 190.2;	DB 3;	Length 5560;
Matches 192;	Conservative	0;	Mismatches 3;	Indels	Gaps
QY	30	TACGTAAGTTCTGCTTCTACCTTGATATATATATAATAATATCATTAATTAGTAGTA	89		
Db	4674	TACGTAAGTTCTGCTTCTACCTTGATATATATATAATAATATCATTAATTAGTAGTA	4615		
QY	90	ATATAATATTTCAAAATATTTTTCAAATAAAAGATGTAGTATATAGCAATTGCTTTT	149		
Db	4614	ATATAATATTTCAAAATATTTTTCAAATAAAAGATGTAGTATATAGCAATTGCTTTT	4555		
QY	150	CTGTAGTTTATAAGTGTGTAATTTTAAATTATAACTTTTCTTAATATATGACCAAAAATTT	209		
Db	4554	CTGTAGTTTATAAGTGTGTAATTTTAAATTATAACTTTTCTTAATATATGACCAAAAACAT	4495		
QY	210	GTGTATGTGCAGGTA	224		
Db	4494	GGTGATGTGCAGGTA	4480		

RESULT 14  
US-09-367-293-2  
; Sequence 2, Application US/09367293  
; Patent No. 6444878  
; GENERAL INFORMATION:  
; APPLICANT: Donaldson, Lain A.  
; APPLICANT: Bøjsen, Kirsten  
; APPLICANT: Jørgensen, Kirsten  
; APPLICANT: Jørsboe, Morten  
; TITLE OF INVENTION: SELECTION METHOD FOR TRANSGENIC PLANTS  
; FILE REFERENCE: DY0U21.001APC  
; CURRENT APPLICATION NUMBER: US/09/367, 293  
; CURRENT FILING DATE: 1999-12-23  
; PRIOR APPLICATION NUMBER: PCT/GB98/00367  
; PRIOR FILING DATE: 1998-02-05  
; PRIOR APPLICATION NUMBER: GB 9702592.8  
; PRIOR FILING DATE: 1997-02-07

```

; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1034
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (0)...(0)
; OTHER INFORMATION: Glucosamine-6-phosphate deaminase.
US-09-367-293-2

```

Query Match	9.4%	Score 190	DB 4	Length 1034
Best Local Similarity	100.0%	Pred. No. 4e-37		
Matches 190	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY 33	GTAAGTTCTGCTTCTACCTTGATATATATATAATTAATTCATTAAATTAGTAGTAATA	92		
Db 235	GTAAGTTCTGCTTCTACCTTGATATATATATAATTAATTCATTAAATTAGTAGTAATA	294		
QY 93	TAATATTTCAAAATATTTTTCACAAATAAAGAATGTAGTATATAGCAATTGCTTTCTG	152		
Db 295	TAATATTTCAAAATATTTTTCACAAATAAAGAATGTAGTATATAGCAATTGCTTTCTG	354		
QY 153	TAGTTTATAGTGTGTATATTTTAAATTTAACTTTTCTTAATATATGACCAAAATTTGTT	212		
Db 355	TAGTTTATAGTGTGTATATTTTAAATTTAACTTTTCTTAATATATGACCAAAATTTGTT	414		
QY 213	GATGTGCAGG 222			
Db 415	GATGTGCAGG 424			

```

RESULT 15
US-09-249-585A-4/c
: Sequence 4, Application US/09249585A
: Patent No. 6417002
:
: GENERAL INFORMATION:
: APPLICANT: Horlick, Robert
: TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
: FILE REFERENCE: 0867/0D905
: CURRENT APPLICATION NUMBER: US/09/249,585A
: CURRENT FILING DATE: 1999-02-11
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 4
: LENGTH: 1926
: TYPE: DNA
: ORGANISM: Epstein Barr Virus
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(1926)
: OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

```

Query Match	2.68;	Score 52.2;	DB 4;	Length 1926;
Best Local Similarity	45.0%;	Pred. No. 0.0022;		
Matches 238;	Conservative 0;	Mismatches 288;	Indels 3;	Gaps 1;
QY 1495	GAAGCGCAGAGAGACGCTTGCGCCTTCTGTGGAAACCCCGACGCTTCGCAAGCTC	1554		
Db 799	GACGGGAGAGACGAGGACGGGAGAGACGGGAGAGACGAGGACGGGAGACGGGAGGAC	740		
QY 1555	ACTGAGCGCCCTGAGAGAGCGCGCAACGGGCGAACCTTGTTCGAGCGCGCCGACGCC	1614		
Db 739	GAGGACGGGAGGAGACGAGGACGGGAGGAGGACGAGGACGGGAGGACGAGGACGGGAGGAC	680		
QY 1615	CTGAACGCCCTTGAAAGAGCTGTACGAAGACCGCGCGCAAGCGGTACGACGGAACCCGTT	1674		
Db 679	GGGGAGGACGGGAGGAGGACGAGGACGGGAGGAGGACGAGGACGGGAGGACGGGAGGAC	620		
QY 1675	GGCAGGAGCACTTCCGGAAGCAACAGGACAGCGCTGACGCTCCGGCAGCAAGGGGCGAA	1734		

```

Db 619 GGGGAGGACGACGACCGGGGAGGAGGACGACGGGGGAGGAGCGGGGAGGAG 560
QY 1735 GAGCGGCTTGCCGAACTT---GAAGCGCCGGAAGCCCCGAACTTCCCTTGACCAATGG 1791
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 559 GACGAGGACGGGAGGAGGACGGGGAGGAGGACGAGCGGGGAGGAGGAGGAGCGGGAG 500
QY 1792 TTCCCGAAGACCGCGACGCTGACCGGACCGGCTTAAGTCGTGTGGGGCGCGCTCA 1851
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 499 GACGGGAGGACGGGAGGAGGACGAGCGGGAGGAGGACGAGCGGAGCGGAGCGGG 440
QY 1852 GTAGACGACGACGCGGTGTCGTGCGGCTCTTGTAGACAGATCGTTGTACAGAGTCG 1911
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 439 GAGGAGGACGAGGACGGGGAGGAGGACGGGGAGGAGCGGAGGAGGAGGAGCGGGAGGAG 380
QY 1912 ACTACGGGACGGGGGAGGAGGACGCCCATCGAGAAAGCGCGCTTCGATCAGTGGCGAAG 1971
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 379 GACGAGGACGGGGAGGAGGACGGGGAGGAGGACGGGGAGGAGGAGGAGGAGGAGCGGGAG 320
QY 1972 CCGCCGACGACGACGAGCAGGAGGAGCGGCCAGGACGGGCGACGGAAGACG 2020
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 319 GAGGACGAGGACGGGGAGGAGGAGGAGCGGGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAG 271
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: May 11, 2003, 23:19:03  
 Job time : 144 secs







Db 181 ATAACCTTTCTAATATATGACCAAAATTTGTGATGTGCGAGGTACGGGGTCTTACGAC 240  
QY 241 CGTCAGTCGGCGGAGCGCGAGAAATTCAGCGCAGCAAGCCCAAGCAGACACCGTAGCGCC 300  
Db 241 CGTCAGTCGGCGGAGCGCGAGAAATTCAGCGCAGCAAGCCCAAGCAGACACCGTAGCGCC 300  
QY 301 AACGAAGACAGCG 360  
Db 301 AACGAAGACAGCG 360  
QY 361 TTCGTGCGGCATTTACGCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420  
Db 361 TTCGTGCGGCATTTACGCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420  
QY 421 GAGTTCGAACGCATCTCGAAGCAATGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
Db 421 GAGTTCGAACGCATCTCGAAGCAATGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
QY 481 GACGTGTGCGGCTTCTCG 540  
Db 481 GACGTGTGCGGCTTCTCG 540  
QY 541 CTGCGCGCTGGCGGTGACGATTTGTTCCACTCAGAAAGCGCTCTCCGCGAGGGAACGTC 600  
Db 541 CTGCGCGCTGGCGGTGACGATTTGTTCCACTCAGAAAGCGCTCTCCGCGAGGGAACGTC 600  
QY 601 ATGACCTGATTCACCTGATTTATGCGGCTCGACGCGTGCACAAAGAAATCTTCGCTGAAG 660  
Db 601 ATGACCTGATTCACCTGATTTATGCGGCTCGACGCGTGCACAAAGAAATCTTCGCTGAAG 660  
QY 661 TCGCGGAAGATTTCTGACACGAAAGAACCTTCAGCGCGGAATTTGGCGGGTACGTGCGCGG 720  
Db 661 TCGCGGAAGATTTCTGACACGAAAGAACCTTCAGCGCGGAATTTGGCGGGTACGTGCGCGG 720  
QY 721 AAGCGCGCTTACGCGCTTCGAGCTTGTTCGAGACGAAAGAGATCACGCGCAACGCGCGA 780  
Db 721 AAGCGCGCTTACGCGCTTCGAGCTTGTTCGAGACGAAAGAGATCACGCGCAACGCGCGA 780  
QY 781 ATGTCATATGTCATCAACAAGCTTGGCGCACTCGACCACTCCCTTACCGGACCGCTTC 840  
Db 781 ATGTCATATGTCATCAACAAGCTTGGCGCACTCGACCACTCCCTTACCGGACCGCTTC 840  
QY 841 GAGTTCGAGCG 900  
Db 841 GAGTTCGAGCG 900  
QY 901 CCCTTCAAGCG 960  
Db 901 CCCTTCAAGCG 960  
QY 961 CGCATGAGCGCTGACCGCTTC 1020  
Db 961 CGCATGAGCGCTGACCGCTTC 1020  
QY 1021 AGCGCTGGGACCGCGCAACCGTTATGCGAATCTTCGGGACCGCGTATTCGGCGCTTC 1080  
Db 1021 AGCGCTGGGACCGCGCAACCGTTATGCGAATCTTCGGGACCGCGTATTCGGCGCTTC 1080  
QY 1081 GCCGTGAGGTGATCTACAAGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140  
Db 1081 GCCGTGAGGTGATCTACAAGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140  
QY 1141 TACCGCATTCAGCG 1200  
Db 1141 TACCGCATTCAGCG 1200  
QY 1201 ATGAGCGCGCTGAGTGTATGAGCTTCAGCGCGTGTGACGCGCAGGCGCGCGCGCGCAAG 1260  
Db 1201 ATGAGCGCGCTGAGTGTATGAGCTTCAGCGCGTGTGACGCGCAGGCGCGCGCGCGCAAG 1260  
QY 1261 GGGCTTCCCGGGGCGAAGCAATTTCTGTCGCGCATGGAAGCTGTACTGCGAGTGTGCG 1320  
Db 1261 GGGCTTCCCGGGGCGAAGCAATTTCTGTCGCGCATGGAAGCTGTACTGCGAGTGTGCG 1320

QY 1321 GCCGTACATGACTTCGAAGCGCGGGGAAGAAATGATCAAGGACTCTTACCGCTGCCGTCGC 1380  
Db 1321 GCCGTACATGACTTCGAAGCGCGGGGAAGAAATGATCAAGGACTCTTACCGCTGCCGTCGC 1380  
QY 1381 CGGAAGGTGATGACCCCGTCCGACCTGGGCGACGACGAAAGGACGTGCAACGTACAGATG 1440  
Db 1381 CGGAAGGTGATGACCCCGTCCGACCTGGGCGACGACGAAAGGACGTGCAACGTACAGATG 1440  
QY 1441 GCGGCATCGACAAGTTCTGTTGCGGAACGCATCTTCAACAAGATCAGGACCGCGGAAGGC 1500  
Db 1441 GCGGCATCGACAAGTTCTGTTGCGGAACGCATCTTCAACAAGATCAGGACCGCGGAAGGC 1500  
QY 1501 GACGAAGAGACGTTGGCGCTTCTGTGGGAAGCGCGCCGACGCTTCGGCAAGCTCAGTAG 1560  
Db 1501 GACGAAGAGACGTTGGCGCTTCTGTGGGAAGCGCGCCCGACGCTTCGGCAAGCTCAGTAG 1560  
QY 1561 GCGCCTGAGAGAGCGCGGCGAACCGCGGAACCTTGTTCGCGAAGCGCGCGCGCGCTGAAC 1620  
Db 1561 GCGCCTGAGAGAGCGCGGCGAACCGCGGAACCTTGTTCGCGAAGCGCGCGCGCGCTGAAC 1620  
QY 1621 GCCCTTGAAGAGCTGTACGAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680  
Db 1621 GCCCTTGAAGAGCTGTACGAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680  
QY 1681 AAGCACTTCGCGGAAGCAACAGCGCGCTGACGCTCCGCGCAGCAAGGCGCGGAAGAGCGG 1740  
Db 1681 AAGCACTTCGCGGAAGCAACAGCGCGCTGACGCTCCGCGCAGCAAGGCGCGGAAGAGCGG 1740  
QY 1741 CTGCGCAACTTGAAGCGCGCGGAAGCGCGGAAGCTTCCCTTGACCAATGTTCCCGGAA 1800  
Db 1741 CTGCGCAACTTGAAGCGCGCGGAAGCGCGGAAGCTTCCCTTGACCAATGTTCCCGGAA 1800  
QY 1801 GACGCGGACGCTGACCG 1860  
Db 1801 GACGCGGACGCTGACCG 1860  
QY 1861 AAGCGGTGTTGTCGCGGCTCTTCTGTAGACAAGATCGTTGTACAGAACTGACTAGCGGC 1920  
Db 1861 AAGCGGTGTTGTCGCGGCTCTTCTGTAGACAAGATCGTTGTACAGAACTGACTAGCGGC 1920  
QY 1921 AGGGGCGAGGGAAGCGCCCATTCGAGAAGCGCGCTTCGATCAGCTGGCGGAAGCGCGGACC 1980  
Db 1921 AGGGGCGAGGGAAGCGCCCATTCGAGAAGCGCGCTTCGATCAGCTGGCGGAAGCGCGGACC 1980  
QY 1981 GACGACGACGAAGACGACGCGCCAGGACGCGCAAGAGAGCTAGCGCGCTAG 2031  
Db 1981 GACGACGACGAAGACGACGCGCCAGGACGCGCAAGAGAGCTAGCGCGCTAG 2031

RESULT 2  
US-09-940-550A-10  
; Sequence 10, Application US/09940550A  
; Publication No. US20030033635A1  
; GENERAL INFORMATION:  
; APPLICANT: MANKIN, LUKE  
; APPLICANT: MCKERSIE, BRYAN  
; TITLE OF INVENTION: SELF-EXCISING POLYNUCLEOTIDES AND USES THEREOF  
; FILE REFERENCE: 16313-0055  
; CURRENT APPLICATION NUMBER: US/09/940,550A  
; CURRENT FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 60/227,961  
; PRIOR FILING DATE: 2000-08-25  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 2031  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: C31int\* DNA  
; OTHER INFORMATION: sequence  
US-09-940-550A-10

Query Match 99.28; Score 2015; DB 9; Length 2031;  
 Best Local Similarity 99.58; Pred. No. 0;  
 Matches 2021; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ATGGCACAAGGGGTGTGACCGGGGGTGATACGTAAGTTCTGCTTCTACCTTGATATA 60  
 Db 1 ATGGCACAAGGGGTGTGACCGGGGGTGATACGTAAGTTCTGCTTCTACCTTGATATA 60  
 QY 61 TAT 120  
 Db 61 TAT 120  
 QY 121 AAGAATAGTATATAGCAATGCTTTCTGTAGTTTATAGTGTATATTTTATTTT 180  
 Db 121 AAGAATAGTATATAGCAATGCTTTCTGTAGTTTATAGTGTATATTTTATTTT 180  
 QY 181 ATAACTTTTCTAATATATGACCAAAATTTGTGTATGTCAGGTACGGGGTCTTACGAC 240  
 Db 181 ATAACTTTTCTAATATATGACCAAAATTTGTGTATGTCAGGTACGGGGTCTTACGAC 240  
 QY 241 CGTCAGTCGGCGGACCGCGAGAAATTCGAGCGGACGACAGCCAGCAGCGTAGCGCC 300  
 Db 241 CGTCAGTCGGCGGACCGCGAGAAATTCGAGCGGACGACGACAGCCAGCAGCGTAGCGCC 300  
 QY 301 AACGAAGACAAGGGCGGCGCCTTCAGCGCGAAGTCGAGCGGAGGGGGCGGTTGAGG 360  
 Db 301 AACGAAGACAAGGGCGGCGCCTTCAGCGCGAAGTCGAGCGGAGGGGGCGGTTGAGG 360  
 QY 361 TTGCTGGGGCATTTTCAGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
 Db 361 TTGCTGGGGCATTTTCAGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
 QY 421 GAGTTCGAACGCATCTGACGAAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
 Db 421 GAGTTCGAACGCATCTGACGAAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
 QY 481 GACGTGCGCGCTTCTCGCGCTGAAGTTCATGAGCGCGATTCGATGTCCTCGGAATTG 540  
 Db 481 GACGTGCGCGCTTCTCGCGCTGAAGTTCATGAGCGCGATTCGATGTCCTCGGAATTG 540  
 QY 541 CTGCGCCCTGGGCGTGACGATGTTTCCACTCAGGAAGCGGCTCTCCGGGAGGAAACGTC 600  
 Db 541 CTGCGCCCTGGGCGTGACGATGTTTCCACTCAGGAAGCGGCTCTCCGGGAGGAAACGTC 600  
 QY 601 ATGACCTGATTCACCTGATTTATGCGGCTGACGCGTCGACAAAGAATCTTCGCTGAAG 660  
 Db 601 ATGACCTGATTCACCTGATTTATGCGGCTGACGCGTCGACAAAGAATCTTCGCTGAAG 660  
 QY 661 TCGCGAAGATTTCTGACACGAAGAACTTCAGCGCGAATTGGCGGTACGTCGGCGGG 720  
 Db 661 TCGCGAAGATTTCTGACACGAAGAACTTCAGCGCGAATTGGCGGTACGTCGGCGGG 720  
 QY 721 AAGCGCCTTACGCTTCGAGCTTGTTCGAGAGCAAGAGATCACGCGCAAGCGCGCA 780  
 Db 721 AAGCGCCTTACGCTTCGAGCTTGTTCGAGAGCAAGAGATCACGCGCAAGCGCGCA 780  
 QY 781 ATGTCATATGTCGTCATCAACAAGCTTGGCGACTCGACCACTCCCTTACCAGACCTTC 840  
 Db 781 ATGTCATATGTCGTCATCAACAAGCTTGGCGACTCGACCACTCCCTTACCAGACCTTC 840  
 QY 841 GAGTTCGAGCCCGAGCGTAATCCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900  
 Db 841 GAGTTCGAGCCCGAGCGTAATCCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900  
 QY 901 CCTTCAGCCGGGAGTCAAGCCCGCATTCACCCGGGAGCATCACGGGGCTTTGTAAG 960  
 Db 901 CCTTCAGCCGGGAGTCAAGCCCGCATTCACCCGGGAGCATCACGGGGCTTTGTAAG 960  
 QY 961 CGCATGAGCGCTGACGCGCTGCCGACCCGGGCGAGAGATTTGGGAAGAACCGCTTCA 1020  
 Db 961 CGCATGAGCGCTGACGCGCTGCCGACCCGGGCGAGAGATTTGGGAAGAACCGCTTCA 1020

QY 1021 AGCGCTGGGACCCGGCAACCGTTATGCGAATCTTCGGGACCCGCGTATTCGGGCTTC 1080  
 Db 1021 AGCGCTGGGACCCGGCAACCGTTATGCGAATCTTCGGGACCCGCGTATTCGGGCTTC 1080  
 QY 1081 GCCGTGAGGTGATCTACAGAAGAAGCCGAGCGGACCGGACCGACGAAGATTGAGGT 1140  
 Db 1081 GCCGTGAGGTGATCTACAGAAGAAGCCGAGCGGACCGGACCGACGAAGATTGAGGT 1140  
 QY 1141 TACCGCATTCAGCGGACCGCATCACGCTCCGGCGGTGAGCTTGATTGCGGACCGATC 1200  
 Db 1141 TACCGCATTCAGCGGACCGCATCACGCTCCGGCGGTGAGCTTGATTGCGGACCGATC 1200  
 QY 1201 ATCGAGCCCGCTGAGTGGTATGAGCTTCAGGCGTGGTTGGACGGGCGGCGCAAG 1260  
 Db 1201 ATCGAGCCCGCTGAGTGGTATGAGCTTCAGGCGTGGTTGGACGGGCGGCGCAAG 1260  
 QY 1261 GGGCTTCCCGGGGGCAAGCCATTCTGTCCGCGCATGACAAAGCTGTACTGCGAGTGGC 1320  
 Db 1261 GGGCTTCCCGGGGGCAAGCCATTCTGTCCGCGCATGACAAAGCTGTACTGCGAGTGGC 1320  
 QY 1321 GCCGTGACTTGAAGCGCGGGAAGAATGATCAAGACTCTTACCCTGCGCTGCGC 1380  
 Db 1321 GCCGTGACTTGAAGCGCGGGAAGAATGATCAAGACTCTTACCCTGCGCTGCGC 1380  
 QY 1381 CGGAAGTGGTGCAGACCCGCTCCGACCTGGGACGACGAGGACAGTGCACATG 1440  
 Db 1381 CGGAAGTGGTGCAGACCCGCTCCGACCTGGGACGACGAGGACAGTGCACATG 1440  
 QY 1441 GCGGCACTGCACAAGTTGCTGTCGGAACGCGATCTTCAACAAGATCAGGACCGCGAAGGC 1500  
 Db 1441 GCGGCACTGCACAAGTTGCTGTCGGAACGCGATCTTCAACAAGATCAGGACCGCGAAGGC 1500  
 QY 1501 GACGAAGAGACGTTGGCGCTTGTGGGAAGCCGCGGACGCTTCGGAAGCTCACTGAG 1560  
 Db 1501 GACGAAGAGACGTTGGCGCTTGTGGGAAGCCGCGGACGCTTCGGAAGCTCACTGAG 1560  
 QY 1561 GCGCTGGAAGAGCGGCGGAAGCGGGAACCTTGTGGGAGCGGCGGCGGCGGCGGCGG 1620  
 Db 1561 GCGCTGGAAGAGCGGCGGAAGCGGGAACCTTGTGGGAGCGGCGGCGGCGGCGGCGG 1620  
 QY 1621 GCCCTGGAAGAGCTGTACGAAGACCGCGGAGCGGCTTACGACGAGCCGCTGGCAGG 1680  
 Db 1621 GCCCTGGAAGAGCTGTACGAAGACCGCGGAGCGGCTTACGACGAGCCGCTGGCAGG 1680  
 QY 1681 AAGCACTTCGGAAGCAACAGGACGCGCTGACGCTCCGCGACGAAGGGCGGAGAAGCGG 1740  
 Db 1681 AAGCACTTCGGAAGCAACAGGACGCGCTGACGCTCCGCGACGAAGGGCGGAGAAGCGG 1740  
 QY 1741 CTTGCCGAACCTGAAGCGCGCGCAAGCCCGGAAGCTTCCCTTGACCAATGTTCCCGAA 1800  
 Db 1741 CTTGCCGAACCTGAAGCGCGCGCAAGCCCGGAAGCTTCCCTTGACCAATGTTCCCGAA 1800  
 QY 1801 GACGCCGAGCTGACCCGAGCCGCTTAAGTGTGGTGGGGGCGGCGTCACTAGACGAC 1860  
 Db 1801 GACGCCGAGCTGACCCGAGCCGCTTAAGTGTGGTGGGGGCGGCGTCACTAGACGAC 1860  
 QY 1861 AAGCGCTGTCTGTCGGGCTTCTGAGCAAGATGTTGTCAAGATGACTAGAGGCGC 1920  
 Db 1861 AAGCGCTGTCTGTCGGGCTTCTGAGCAAGATGTTGTCAAGATGACTAGAGGCGC 1920  
 QY 1921 AGGGGCAAGGAAGCCCATGAGAAGCGGCTTGCATCAGTGGGGAAGCCGCGGAC 1980  
 Db 1921 AGGGGCAAGGAAGCCCATGAGAAGCGGCTTGCATCAGTGGGGAAGCCGCGGAC 1980  
 QY 1981 GACGACGACGAAGAGAGAGCGCCGACGAGCGGACGGAAGACTAGCGGCTAG 2031  
 Db 1981 GACGACGACGAAGAGAGAGCGCCGACGAGCGGACGGAAGACTAGCGGCTAG 2031

RESULT 3  
 US-09-940-550A-11  
 ; Sequence 11, Application US/09940550A  
 ; Publication No. US20030033635A1

```

: GENERAL INFORMATION:
: APPLICANT: MANKIN, LUKE
: APPLICANT: MCKERSIE, BRYAN
: TITLE OF INVENTION: SELF-EXCISING POLYNUCLEOTIDES AND USES THEREOF
: FILE REFERENCE: 16313-0055
: CURRENT APPLICATION NUMBER: US/09/940,550A
: PRIOR FILING DATE: 2001-08-27
: PRIOR APPLICATION NUMBER: 60/227,961
: PRIOR FILING DATE: 2000-08-25
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 11
: LENGTH: 9901
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: pBps EW051
: OTHER INFORMATION: T-DNA region
: NAME/KEY: modified_base
: LOCATION: (1984)
: OTHER INFORMATION: a, t, c, g, other or unknown
: NAME/KEY: modified_base
: LOCATION: (1986)
: OTHER INFORMATION: a, t, c, g, other or unknown
: NAME/KEY: modified_base
: LOCATION: (5231)
: OTHER INFORMATION: a, t, c, g, other or unknown
: NAME/KEY: modified_base
: LOCATION: (5233)
: OTHER INFORMATION: a, t, c, g, other or unknown
: NAME/KEY: modified_base
: LOCATION: (8478)
: OTHER INFORMATION: a, t, c, g, other or unknown
: NAME/KEY: modified_base
: LOCATION: (8480)
: OTHER INFORMATION: a, t, c, g, other or unknown
: US-09-940-550A-11

```

Query Match 99.2%; Score 2014; DB 9; Length 9901;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2031; Conservative 0; Mismatches 0; Indels 7; Gaps 1;

```

QY 1 ATGGCACAAGGGGTGTGACCGGGGTGATACGTAAGTTTCTTCTACCTTGATATA 60
Db 4984 ATGGCACAAGGGGTGTGACCGGGGTGATACGTAAGTTTCTTCTACCTTGATATA 5043
QY 61 TATATAATAATATCATTAATTAGTAATAATAATATTAATTTTTCAAAATA 120
Db 5044 TATATAATAATATCATTAATTAGTAATAATAATATTAATTTTTCAAAATA 5103
QY 121 AAAGAATGTAGTATATAGCAATTCCTTCTGTAGTTTAAAGTGTATATTTTAATT 180
Db 5104 AAAGAATGTAGTATATAGCAATTCCTTCTGTAGTTTAAAGTGTATATTTTAATT 5163
QY 181 ATAACTTTTCTAATATATGACCAAAATTTGTTGATGTGACGTACGGGGTCTTACGAC 240
Db 5164 ATAACTTTTCTAATATATGACCAAAATTTGTTGATGTGACGTACGGGGTCTTACGAC 5223
QY 241 CGTC-----AGTCGGCGAGCGCGAGAAATTCGAGCGCAGCAAGCCCGACACAGCG 293
Db 5224 CGTCGRCNTNDAGTCGCGGAGCGGAGAAATTCGAGCGCAGCAAGCCCGACACAGCG 5283
QY 294 TAGCGCCAAGAGACAGAGCGCGCCGACCTTCAGCGCGAAGTGCAGCGGAGCGGGGCGG 353
Db 5284 TAGCGCCAAGAGACAGAGCGCGCCGACCTTCAGCGCGAAGTGCAGCGGAGCGGGGCGG 5343
QY 354 GTTCAGGTTCTGCGGCATTTACGGAAGCGCGCGCGCGCGCTTCGCGGAGCGCGGA 413
Db 5344 GTTCAGGTTCTGCGGCATTTACGGAAGCGCGCGCGCGCGCTTCGCGGAGCGCGGA 5403
QY 414 GCGCCCGGAGTTCGAACGATCTGAACGAATGCGCGCGCGCGGCGCTCAACATGATCAT 473
Db 5404 GCGCCCGGAGTTCGAACGATCTGAACGAATGCGCGCGCGCGGCGCTCAACATGATCAT 5463

```

```

QY 474 TGCTATGACGTTCGCGCTTCTCGCCCTGAAGGTCATGACGCGATTCGGATTGCTC 533
Db 5464 TGCTATGACGTTCGCGCTTCTCGCCCTGAAGGTCATGACGCGATTCGGATTGCTC 5523
QY 534 GGAATTGCTCGCCCTGCGGTGACGATGTTTCCACTCAGGAAGCGCTTCCGCGAGGG 593
Db 5524 GGAATTGCTCGCCCTGCGGTGACGATGTTTCCACTCAGGAAGCGCTTCCGCGAGGG 5583
QY 594 AACGTCATGACCTGATTCACCTGATTTATGCGGCTCGACCGCTCGACAAAGATCTTC 653
Db 5584 AACGTCATGACCTGATTCACCTGATTTATGCGGCTCGACCGCTCGACAAAGATCTTC 5643
QY 654 GCTGAAGTCGCGGAAGATTCTCGACACGAAGAACCCTTCAGCGGAATTGGCGGCTACGT 713
Db 5644 GCTGAAGTCGCGGAAGATTCTCGACACGAAGAACCCTTCAGCGGAATTGGCGGCTACGT 5703
QY 714 CGCGGGAAGCGCCCTTACGCGTTGAGCTTGTTCGGAAGCAAGAGATCAGCGCAA 773
Db 5704 CGCGGGAAGCGCCCTTACGCGTTGAGCTTGTTCGGAAGCAAGAGATCAGCGCAA 5763
QY 774 CGCGGGAATGTCATATGTCGTATCAACAAGCTTGGCGCACTCGACCACTCCCTTACCGG 833
Db 5764 CGCGGGAATGTCATATGTCGTATCAACAAGCTTGGCGCACTCGACCACTCCCTTACCGG 5823
QY 834 ACCCTTCGAGTTCGAGCCCGACGTAATCCGGTGTGTGCGGTGAGATCAAGACGCACAA 893
Db 5824 ACCCTTCGAGTTCGAGCCCGACGTAATCCGGTGTGTGCGGTGAGATCAAGACGCACAA 5883
QY 894 ACACCTTCCTTCAAGCCCGGCGAGTCAAGCCCGCATTCACCCGGCAGCATCACGGGGCT 953
Db 5884 ACACCTTCCTTCAAGCCCGGCGAGTCAAGCCCGCATTCACCCGGCAGCATCACGGGGCT 5943
QY 954 TTGTAAGCGCATGAGCGCTGACCGCGTGCAGCCCGGGCGAGACGATGGGAAGAAGAC 1013
Db 5944 TTGTAAGCGCATGAGCGCTGACCGCGTGCAGCCCGGGCGAGACGATGGGAAGAAGAC 6003
QY 1014 CGCTTCAAGCGCCTGGGACCGGCAACCGTTATGCGAATCCTTCGGGACCCGCTATTGC 1073
Db 6004 CGCTTCAAGCGCCTGGGACCGGCAACCGTTATGCGAATCCTTCGGGACCCGCTATTGC 6063
QY 1074 GGGCTTCGCGCTGAGGTGATCTTCAAGAAGAAGCCGAGCGGACGCGGACGACGAGAT 1133
Db 6064 GGGCTTCGCGCTGAGGTGATCTTCAAGAAGAAGCCGAGCGGACGCGGACGACGAGAT 6123
QY 1134 TGAGGGTTACCGCATTCAGCGCGGACCCGATCAAGCTCCGCGGTGAGCTTGATTGCGG 1193
Db 6124 TGAGGGTTACCGCATTCAGCGCGGACCCGATCAAGCTCCGCGGTGAGCTTGATTGCGG 6183
QY 1194 ACCGATCATGAGCCCGCTGAGTGTATGACCTTCAGGCGTGTGTGACGCGGAGGGGCG 1253
Db 6184 ACCGATCATGAGCCCGCTGAGTGTATGACCTTCAGGCGTGTGTGACGCGGAGGGGCG 6243
QY 1254 CGGCAAGGGGCTTCCCGGGGGCAAGCCATTTCTGCCCATGGAAGCTGTACTGCGA 1313
Db 6244 CGGCAAGGGGCTTCCCGGGGGCAAGCCATTTCTGCCCATGGAAGCTGTACTGCGA 6303
QY 1314 GTGTGGCGCGCTCATGACTTCGAAGCGCGGGGGAAGATGATCAAGGACTTTACCGCTG 1373
Db 6304 GTGTGGCGCGCTCATGACTTCGAAGCGCGGGGGAAGATGATCAAGGACTTTACCGCTG 6363
QY 1374 CCGTCGCGGGAAGGTGTGACCCGCTCGCACTTGGGCGAGCAGAAAGCAGTGAACGT 1433
Db 6364 CCGTCGCGGGAAGGTGTGACCCGCTCGCACTTGGGCGAGCAGAAAGCAGTGAACGT 6423
QY 1434 CAGCATGGCGGCACTGACAAGTTGTTGCGGAAGCATCTTCAACAAGATCAGGCAAGC 1493
Db 6424 CAGCATGGCGGCACTGACAAGTTGTTGCGGAAGCATCTTCAACAAGATCAGGCAAGC 6483
QY 1494 CGAAGCGAGAGAGACGTTGGCGTTCTGTGGAGCGCGCGGAGCGCTTCGCGCAAGCT 1553
Db 6484 CGAAGCGAGAGAGACGTTGGCGTTCTGTGGAGCGCGCGGAGCGCGCTTCGCGCAAGCT 6543

```

OY 1554 CACTGAGCGCCTGAGAGAGCGGCGAAGCGGCGAACTTGTTCGGAGCGCGCGCAGCGC 1613  
|||||  
Db 6544 CACTGAGCGCCTGAGAGAGCGGCGAAGCGGCGAACTTGTTCGGAGCGCGCGCAGCGC 6603  
OY 1614 CCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGCAGCGCGCTACGACCGACCCGT 1673  
|||||  
Db 6604 CCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGCAGCGCGCTACGACCGACCCGT 6663  
OY 1674 TGGCAGGAAGACACTTCCGGAAGCAACAGGCGACGCGCTCCGGCAGCAAGGGCGGA 1733  
|||||  
Db 6664 TGGCAGGAAGACACTTCCGGAAGCAACAGGCGACGCGCTCCGGCAGCAAGGGCGGA 6723  
OY 1734 AGAGCGGCTTGCCTGAACTTGAAGCGCGCGAAGCGCGGAACTTCCCTTGAACCAATGGTT 1793  
|||||  
Db 6724 AGAGCGGCTTGCCTGAACTTGAAGCGCGCGAAGCGCGGAACTTCCCTTGAACCAATGGTT 6783  
OY 1794 CCCCCAAGACGCGGACGCTGACCCGACCGCGCCTTAAGTCTGTGTGGGGCGCGCTCAGT 1853  
|||||  
Db 6784 CCCCCAAGACGCGGACGCTGACCCGACCGCGCCTTAAGTCTGTGTGGGGCGCGCTCAGT 6843  
OY 1854 AGACGACAAGCGCGTGTCTGTCGGGCTCTTCGTAGACAAGATCGTTGCACGAAGTCGAC 1913  
|||||  
Db 6844 AGACGACAAGCGCGTGTCTGTCGGGCTCTTCGTAGACAAGATCGTTGCACGAAGTCGAC 6903  
OY 1914 TACGGGCGAGGGGCGAGGGAAGCGCCCATCGAGAAGCGCGCTTCGATCAGTGGGCGAAGCC 1973  
|||||  
Db 6904 TACGGGCGAGGGGCGAGGGAAGCGCCCATCGAGAAGCGCGCTTCGATCAGTGGGCGAAGCC 6963  
OY 1974 GCCGACCGGACGACGACGAGAAGACGACGCCGACGAGCGGACGGAAGACGTACGGCGTAG 2031  
|||||  
Db 6964 GCCGACCGGACGACGACGAGAAGACGACGCCGACGAGCGGACGGAAGACGTACGGCGTAG 7021

RESULT 4  
US-09-788-297-20

; Sequence 20, Application US/09788297  
; Patent No. US20020094516A1  
; GENERAL INFORMATION:  
; APPLICANT: Calos, Michele P.  
; APPLICANT: Scilment, Christopher R.  
; TITLE OF INVENTION: ALTERED RECOMBINASES FOR GENOME MODIFICATION  
; FILE REFERENCE: 8400-0011  
; CURRENT APPLICATION NUMBER: US/09/788,297  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 1842  
; TYPE: DNA  
; ORGANISM: Phage phiC31  
US-09-788-297-20

Query Match 89.2%; Score 1811; DB 10; Length 1842;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1814; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
OY 213 GATGTGACGTACGCGGCTTACGACCGTACGTCGCGGAGCGGAGAAATTCGAGCGC 272  
| | | | |  
Db 24 GGTGACACGTACGCGGCTTACGACCGTACGTCGCGGAGCGGAAATTCGAGCGC 83  
OY 273 AGCAAGCCGACGACACACAGCGTAGCGCAACGAGAAGAGGCGCGACCTTCAGCGCA 332  
| | | | |  
Db 84 AGCAAGCCGACGACACACAGCGTAGCGCAACGAGAAGGCGCGACCTTCAGCGCA 143  
OY 333 AGTCGAGCGGAGCGGCGGCTTACGAGTTCGTCGGCATTTCAAGCGAAGCGCGGCGAC 392  
| | | | |  
Db 144 AGTCGAGCGGAGCGGCGGCTTACGAGTTCGTCGGCATTTCAAGCGAAGCGCGGCGAC 203  
OY 393 GTCGGCGTTGCGGAGCGGCGGAGCGCGGAGTTCGAACGCATCTGAACGAATGCGCGC 452  
| | | | |  
Db 204 GTCGGCGTTGCGGAGCGGCGGAGCGCGGAGTTCGAACGCATCTGAACGAATGCGCGC 263  
OY 453 CGGGCGGCTCAACATGATCATTTGTATGACGTCTCGCGCTTCTCGCGCTGAAGTCAT 512

|||||  
Db 264 CGGGCGGCTCAACATGATCATTTGTCTATGACGTTCGCGCTTCTCGCGCTGAAGTCAT 323  
OY 513 GAGCGGATTCGATTTGTCTCGGAATTGCTCGCCCTGGGCGTGACGATTTCTTCCACTCA 572  
|||||  
Db 324 GAGCGGATTCGATTTGTCTCGGAATTGCTCGCCCTGGGCGTGACGATTTCTTCCACTCA 383  
OY 573 GGAAGCGCTTCCGCGGAGGAAACGTATGACCTGATTCACCTGATTAATGCGGCTCGA 632  
| | | | |  
Db 384 GGAAGCGCTTCCGCGGAGGAAACGTATGACCTGATTCACCTGATTAATGCGGCTCGA 443  
OY 633 CGCGTCGACAAAGAAATCTTCGTAAGTCCGGCGAAGATTCGACACGAAAGAACTTCA 692  
| | | | |  
Db 444 CGCGTCGACAAAGAAATCTTCGTAAGTCCGGCGAAGATTCGACACGAAAGAACTTCA 503  
OY 693 GCGCAATTGGCGGCTACGTCGCGGCGGGAAGGCGCTTACGCGCTTCGAGCTTGTTCGA 752  
| | | | |  
Db 504 GCGCAATTGGCGGCTACGTCGCGGCGGGAAGGCGCTTACGCGCTTCGAGCTTGTTCGA 563  
OY 753 GACGAAGAGATCAGCGCGCAACGCGGCAATGTCATATGTCATCAACAGCTTGCGCA 812  
| | | | |  
Db 564 GACGAAGAGATCAGCGCGCAACGCGGCAATGTCATATGTCATCAACAGCTTGCGCA 623  
OY 813 CTGCAACACTCCCTTACCGGACCTTCGAGTTGAGCCCGGACGTAATCCGCTGTGTG 872  
| | | | |  
Db 624 CTGCAACACTCCCTTACCGGACCTTCGAGTTGAGCCCGGACGTAATCCGCTGTGTG 683  
OY 873 GCGTGAGATCAAGACGACAAACACTTCCCTTCAAGCCCGGCGAGTCAAGCGCCATTCA 932  
| | | | |  
Db 684 GCGTGAGATCAAGACGACAAACACTTCCCTTCAAGCCCGGCGAGTCAAGCGCCATTCA 743  
OY 933 CCGGCGACGATCAGCGGCGCTTGTAAAGCGCATGACGCTGACGCCGTGCGGACCCGGG 992  
| | | | |  
Db 744 CCGGCGACGATCAGCGGCGCTTGTAAAGCGCATGACGCTGACGCCGTGCGGACCCGGG 803  
OY 993 CGAGCAGATTGGGAAGAGAAGCCGCTTCAAGCGCCTGGGAACCCGCAACCGTTATCGAAT 1052  
| | | | |  
Db 804 CGAGCAGATTGGGAAGAGAAGCCGCTTCAAGCGCCTGGGAACCCGCAACCGTTATCGAAT 863  
OY 1053 CCTTCGGGACCCGCGTATTCGCGGCTTCGCCGCTGAGTGTATCTACAAGAGAAGCCGGA 1112  
| | | | |  
Db 864 CCTTCGGGACCCGCGTATTCGCGGCTTCGCCGCTGAGTGTATCTACAAGAGAAGCCGGA 923  
OY 1113 CGGCAAGCGGACGACGAGAAGATTGAGGTTACCGCATTCAGCGGACCGCATCAGCTCCG 1172  
| | | | |  
Db 924 CGGCAAGCGGACGACGAGAAGATTGAGGTTACCGCATTCAGCGGACCGCATCAGCTCCG 983  
OY 1173 GCGGTCGAGCTTGATTCGCGGACGATCATCGACCCGCTGAGTGTATGAGCTTCAGGC 1232  
| | | | |  
Db 984 GCGGTCGAGCTTGATTCGCGGACGATCATCGACCCGCTGAGTGTATGAGCTTCAGGC 1043  
OY 1233 GTGTTGACGCGGACGAGGCGCGGCAAGGCGCTTCCCGGGGGCAAGCCATCTGTCCGC 1292  
| | | | |  
Db 1044 GTGTTGACGCGGACGAGGCGCGGCAAGGCGCTTCCCGGGGGCAAGCCATCTGTCCGC 1103  
OY 1293 CATGACAAGCTGTACTGCGAGTGTGGCGGCTCATGACTTCGAAGCGCGGGAAGAATC 1352  
| | | | |  
Db 1104 CATGACAAGCTGTACTGCGAGTGTGGCGGCTCATGACTTCGAAGCGCGGGAAGAATC 1163  
OY 1353 GATCAAGGACTCTTACCGCTGCGGCGGGAAGGTGTCGACCCGTCGCACTTGCGCA 1412  
| | | | |  
Db 1164 GATCAAGGACTCTTACCGCTGCGGCGGGAAGGTGTCGACCCGTCGCACTTGCGCA 1223  
OY 1413 GCACGAAGGACGTCGAACGTCAGCATGGCGGACATGCAAGTTGCTGCGGAACGCAAT 1472  
| | | | |  
Db 1224 GCACGAAGGACGTCGAACGTCAGCATGGCGGACATGCAAGTTGCTGCGGAACGCAAT 1283  
OY 1473 CTTCAACAAGATCAGGCAAGCGGCAAGGCGGACGAGAGACGTTGGCGCTTCTGTGGAAGC 1532  
| | | | |  
Db 1284 CTTCAACAAGATCAGGCAAGCGGCAAGGCGGACGAGAGACGTTGGCGCTTCTGTGGAAGC 1343  
OY 1533 CGCCGACGCTTCGGCAAGCTTCATGAGGCGCTGAGAGAAGCGGCGAAGCGGCAACCT 1592  
| | | | |



Db 1344 CGCCCGACGCTTCGCGCAAGCTCACTGAGCGCCTGAGAGAGCGCGAACCAGGCGCAACT 1403  
QY 1593 TGTTCGGAGCGCGCGCGACGCCCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGC 1652  
Db 1404 TGTTCGGAGCGCGCGCGACGCCCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGC 1463  
QY 1653 AGCGCGGTACGACGAGCGCGCTTGGCAGGAGCACTTCCGGAAGCAAGAGCAGCGCTGAC 1712  
Db 1464 AGCGCGGTACGACGAGCGCGCTTGGCAGGAGCACTTCCGGAAGCAAGAGCAGCGCTGAC 1523  
QY 1713 GCTCCGGCAGCAAGGGGGCGGAGAGCGGCTTGGCAACTTGAAGCGCGCGAAGCGCGAA 1772  
Db 1524 GCTCCGGCAGCAAGGGGGCGGAGAGCGGCTTGGCAACTTGAAGCGCGCGAAGCGCGAA 1583  
QY 1773 GCTTCCCTTGACCAATGTTCCCGAAGACCGCGAGCTGACCGCGCGCTTAAGTC 1832  
-Db 1584 GCTTCCCTTGACCAATGTTCCCGAAGACCGCGAGCTGACCGCGCGCTTAAGTC 1643  
QY 1833 GTGTTGGGGCGCGCTCAGTAGACGACAAAGCGGTGTTGTCGGGCTCTTGTAGACAA 1892  
Db 1644 GTGTTGGGGCGCGCTCAGTAGACGACAAAGCGGTGTTGTCGGGCTCTTGTAGACAA 1703  
QY 1893 GATCGTTGTACGAAGTCACTACGAGGCGAGGGGCGAGGAGCAAGCGCCATCGAGACCGCGC 1952  
Db 1704 GATCGTTGTACGAAGTCACTACGAGGCGAGGGGCGAGGAGCAAGCGCCATCGAGACCGCGC 1763  
QY 1953 TTGATCAGTGGGGCGAAGCGCGCGACCGACGACGACGACGACGACGACGACGACGACGAC 2012  
Db 1764 TTGATCAGTGGGGCGAAGCGCGCGACCGACGACGACGACGACGACGACGACGACGACGAC 1823  
QY 2013 GGAAGACGTAGCGGCGTAG 2031  
Db 1824 GGAAGACGTAGCGGCGTAG 1842

RESULT 5  
US-09-788-297-26  
; Sequence 26, Application US/09788297  
; Patent No. US20020094516A1  
; GENERAL INFORMATION:  
; APPLICANT: Calos, Michele P.  
; APPLICANT: Scilmenti, Christopher R.  
; TITLE OF INVENTION: ALTERED RECOMBINASES FOR GENOME MODIFICATION  
; FILE REFERENCE: 8400-0011  
; CURRENT APPLICATION NUMBER: US/09/788,297  
; CURRENT FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 26  
; LENGTH: 1842  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: altered  
; OTHER INFORMATION: recombinase 7C1  
US-09-788-297-26

Query Match 88.9%; Score 1806.2; DB 10; Length 1842;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1811; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 213 GATGTGCAGGTACCGGGTCTTACGACCGTCAAGTCGCGGAGCGGAGAGAAATTCAGCGC 272  
Db 24 GTTGACACAGTACCGGGTCTTACGACCGTCAAGTCGCGGAGCGGAGAAATTCAGCGC 83  
QY 273 AGCAAGCCCGACGACACAGCGTAGCGCAAGCAAGCAAGCGCGCGACCTTCAGCGCGA 332  
Db 84 AGCAAGCCCGACGACACAGCGTAGCGCAAGCAAGCAAGCGCGCGACCTTCAGCGCGA 143  
QY 333 AGTCAGCGCGACGCGGGCGGCTTCAAGTTCGTCGGGCAATTCAGCGAAGCGCGCGGAC 392  
Db 144 AGTCAGCGCGACGCGGGCGGCTTCAAGTTCGTCGGGCAATTCAGCGAAGCGCGCGGAC 203

QY 393 GTCGGCTTCGGGACGGCGGAGCGCCCGGAGTTGCAAGCATCTGAAACGAATGCCGCGC 452  
Db 204 GTCGGCTTCGGGACGGCGGAGCGCCCGGAGTTGCAAGCATCTGAAACGAATGCCGCGC 263  
QY 453 CGGGCGCTCAACATGATCATTTCTATGACGTGTGCGGCTTCTCGCGCGCTGAAGTCAAT 512  
Db 264 CGGGCGCTCAACATGATCATTTCTATGACGTGTGCGGCTTCTCGCGCGCTGAAGTCAAT 323  
QY 513 GGACCGCATTCGATTTGTCGGAATTTGCTCGCCCTGGGCGTGACGATTTTCCACTCA 572  
Db 324 GGACCGCATTCGATTTGTCGGAATTTGCTCGCCCTGGGCGTGACGATTTTCCACTCA 383  
QY 573 GGAAGCGCTCTTCGGCAGGGAACGTCAATGACCTGATTCACCTGATTAATGCGCTCA 632  
Db 384 GGAAGCGCTCTTCGGCAGGGAACGTCAATGACCTGATTCACCTGATTAATGCGCTCA 443  
QY 633 CGCGTCGACAAAGAAATCTTCGCTGAAGTCGGCGAAGATTTCTGACAGCAAGACCTTCA 692  
Db 444 CGCGTCGACAAAGAAATCTTCGCTGAAGTCGGCGAAGATTTCTGACAGCAAGACCTTCA 503  
QY 693 GCGCAATTTGGCGGGTACGTGCGGCGGGAAGCGCCTTACGGCTTTCGAGCTTGTTCGA 752  
Db 504 GCGCAATTTGGCGGGTACGTGCGGCGGGAAGCGCCTTACGGCTTTCGAGCTTGTTCGA 563  
QY 753 GACGAAGAGATCACGCGCAAGCGCGCAATGTCAATGTCTCATCAACAAGCTTGCCTCA 812  
Db 564 GACGAAGAGATCACGCGCAAGCGCGCAATGTCAATGTCTCATCAACAAGCTTGCCTCA 623  
QY 813 CTCGACCACTCCCTTACCGGAGCCCTTGAGTTTCGAGCCCGACGTAATCCGCTGTGCTG 872  
Db 624 CTCGACCACTCCCTTACCGGAGCCCTTGAGTTTCGAGCCCGACGTAATCCGCTGTGCTG 683  
QY 873 GCGTAGATCAAGACGACCAAAACACCTTCCCTTCAAGCCCGGACGTAAGCCGCATTTCA 932  
Db 684 GCGTAGATCAAGACGACCAAAACACCTTCCCTTCAAGCCCGGACGTAAGCCGCATTTCA 743  
QY 933 CCGGGGACGATCACGGGGCTTGTAAAGCCCATGAGCGCTGACCGCGCGCGCGCGCGG 992  
Db 744 CCGGGGACGATCACGGGGCTTGTAAAGCCCATGAGCGCTGACCGCGCGCGCGCGCGG 803  
QY 993 CGAGACGATTTGGGAAGAAGACCGCTTCAAGCGCCTGGGACCGCGCAACCGTTATGCCAAT 1052  
Db 804 CGAGACGATTTGGGAAGAAGACCGCTTCAAGCGCCTGGGACCGCGCAACCGTTATGCCAAT 863  
QY 1053 CCTTCGGGACCGCGTATTTGGGGCTTCCCGCTGAGGTGATCTCAAGAAGAAGCGGGA 1112  
Db 864 CCTTCGGGACCGCGTATTTGGGGCTTCCCGCTGAGGTGATCTCAAGAAGAAGCGGGA 923  
QY 1113 CGGCACGCGGACCAAGAGATTGAGGTTACCGCATTCAGCGCGCAACCGATCACGCTCCG 1172  
Db 924 CGGCACGCGGACCAAGAGATTGAGGTTACCGCATTCAGCGCGCAACCGATCACGCTCCG 983  
QY 1173 GCCGGTCGAGCTTGAATTCGGGACCGATCATGACCGCGCTGAGTGTATGAGCTTCAGGC 1232  
Db 984 GCCGGTCGAGCTTGAATTCGGGACCGATCATGACCGCGCTGAGTGTATGAGCTTCAGGC 1043  
QY 1233 GTGTTGACGCGCAGGGGGCGCGCAAGGGGCTTTCGCGGCGCAAGCCATTTCTGCCG 1292  
Db 1044 GTGTTGACGCGCAGGGGGCGCGCAAGGGGCTTTCGCGGCGCAAGCCATTTCTGCCG 1103  
QY 1293 CATGACAAGCTGTACTGCGAGTGTGCGCGCTCATGACTTCGAAGCGCGGGAAGAATC 1352  
Db 1104 CATGACAAGCTGTACTGCGAGTGTGCGCGCTCATGACTTCGAAGCGCGGGAAGAATC 1163  
QY 1353 GATCAAGGACTTTACCGCTGCCGTGCGCGGAAGGTGTGACCGCTCCGACCTGGGCA 1412  
Db 1164 GATCAAGGACTTTACCGCTGCCGTGCGCGGAAGGTGTGACCGCTCCGACCTGGGCA 1223  
QY 1413 GCACGAAGCAGCTGCAACGTACATGCGGCGCTGACCAAGTTCGTTGCGGAACGCAT 1472  
Db 1224 GCACGAAGCAGCTGCAACGTACATGCGGCGCTGACCAAGTTCGTTGCGGAACGCAT 1283  
QY 1473 CTTCAACAAGATCAGGACCGCGAAGCGGACGAGAGACGTTGGCGCTTCTGTGGGAAGC 1532



Db 1284 CTTCAACAAGATCAGGACGCCGAAGCGACGACGAGACGTTGGCGCTTCTGTGGGAAGC 1343  
QY 1533 CGCCCGACGCTTCGGCAAGCTCACTGAGGCGGCTGAGAAGAGCGGCGGAACCGGAACCT 1592  
Db 1344 CGCCCGACGCTTCGGCAAGCTCACTGAGGCGGCTGAGAAGAGCGGCGGAACCGGAACCT 1403  
QY 1593 TGTTCGGAGCGCGCCGACGCCCTGAAACGCCCTTGAAGAGCTGTACGAGACCGCGCGGC 1652  
Db 1404 TGTTCGGAGCGCGCCGACGCCCTGAAACGCCCTTGAAGAGCTGTACGAGACCGCGCGGC 1463  
QY 1653 AGCGCGCTACGAGCGAGCCGTTGGCAGGAAGCACTTCGGGAAGCAACAGCAGCGCTGAC 1712  
Db 1464 AGCGCGCTACGAGCGAGCCGTTGGCAGGAAGCACTTCGGGAAGCAACAGCAGCGCTGAC 1523  
QY 1713 GCTCCGGCAGCAAGGCGCGGAAGAGCGGCTTGGCCGAACCTTGAAGCGCGCGGAAGCCCGAA 1772  
Db 1524 GCTCCGGCAGCAAGGCGCGGAAGAGCGGCTTGGCCGAACCTTGAAGCGCGCGGAAGCCCGAA 1583  
QY 1773 GCTTCCCTTGACCAATGTTCCCGGAAGACGCCGACGCTGACCCGACCGCCCTAAGTC 1832  
Db 1584 GCTTCCCTTGACCAATGTTCCCGGAAGACGCCGACGCTGACCCGACCGCCCTAAGTC 1643  
QY 1833 GTGTGGGGGGCGCGCTACGTAGACGACGACGCGCTGCTGCGGGCTTTCTGTAGACAA 1892  
Db 1644 GTGTGGGGGGCGCGCTACGTAGACGACGACGCGCTGCTGCGGGCTTTCTGTAGACAA 1703  
QY 1893 GATCGTTGTACGAAGTCACTACGGGCGAGGGGCGGAAGCGCCCATCGAAGCGCGC 1952  
Db 1704 GATCGTTGTACGAAGTCACTACGGGCGAGGGGCGGAAGCGCCCATCGAAGCGCGC 1763  
QY 1953 TTTCATCAGTGGGCGAAGCGCGCGACCGACGACGACGAGCAAGACGCGCCGACGCGCAC 2012  
Db 1764 TTTCATCAGTGGGCGAAGCGCGCGACCGACGACGACGAGCAAGACGACGCGCCGACGCGCAC 1823  
QY 2013 GGAAGACGTAGCGCGCTAG 2031  
Db 1824 GCAAGACGTAGCGCGCTAG 1842

RESULT 6

US-09-788-297-27

; Sequence 27, Application US/09788297  
; Patent No. US20020094516A1  
; GENERAL INFORMATION:  
; APPLICANT: Calos, Michele P.  
; APPLICANT: Scilmenti, Christopher R.  
; TITLE OF INVENTION: ALTERED RECOMBINASES FOR GENOME MODIFICATION  
; FILE REFERENCE: 8400-0011  
; CURRENT APPLICATION NUMBER: US/09/788,297  
; CURRENT FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 1839  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: altered  
; OTHER INFORMATION: recombinase 5C1  
US-09-788-297-27

Query Match 88.7%; Score 1801.6; DB 10; Length 1839;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1807; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 213 GATGTGCAGTAGCGGGGTCTTACGACCGTCACTGCGCGGAGCGGAGAAATTCGAGCGC 272  
Db 24 GGTGCACACGTACGCGGGTCTTACGACCGTCACTGCGCGGAGCGGAGAAATTCGAGCGC 83  
QY 273 AGCAAGCCGACGACACAGCGTAGCGGCAAGCAAGAGCGGCGCGACCTTCAGCGCGA 332  
Db 84 AGCAAGCCGACGACACAGCGTAGCGGCAAGCAAGAGCGGCGCGACCTTCAGCGCGA 143

QY 333 AGTCAGCGCGACGCGGGCGGCTTACAGTTCTGTGGGCAATTCAGCGAAGCGCGGCGAC 392  
Db 144 AGTCAGCGCGACGCGGGCGGCTTACAGTTCTGTGGGCAATTCAGCGAAGCGCGGCGAC 203  
QY 393 GTCGGGCTTCGGGACGCGCGGAGCGCGCGGAGTTTCGAACGCATCCTGAACGAATGCGCGC 452  
Db 204 GTCGGGCTTCGGGACGCGCGGAGCGCGCGGAGTTTCGAACGCATCCTGAACGAATGCGCGC 263  
QY 453 CGGGCGCTCAACATGATCATGTGTATGACGTGTGCGGCTTCTGCGCGCTGAAGGTCAT 512  
Db 264 CGGGCGCTCAACATGATCATGTGTATGACGTGTGCGGCTTCTGCGCGCTGAAGGTCAT 323  
QY 513 GGACGCAATTCGATTTGTCTCGAATTTGCTCGCCCTGGGCGTGAAGATTTGTTCCACTCA 572  
Db 324 GGACGCAATTCGATTTGTCTCGAATTTGCTCGCCCTGGGCGTGAAGATTTGTTCCACTCA 383  
QY 573 GGAAGGCGTCTTCGGGAGGAAACGTATGACCTGATTAAGCGGCTCGA 632  
Db 384 GGAAGGCGTCTTCGGGAGGAAACGTATGACCTGATTAAGCGGCTCGA 443  
QY 633 CGCGTCCACAAAGATCTTCCGTGAAGTCCGGGCAAGATTTCTGACACGAAGAACCTTCA 692  
Db 444 CGCGTCCACAAAGATCTTCCGTGAAGTCCGGGCAAGATTTCTGACACGAAGAACCTTCA 503  
QY 693 GCGCGAATTTGGGCGGGTACGTCGCGCGGGAAGGCGGCTTACGCGCTTCGAGCTTGTTCGGA 752  
Db 504 GCGCGAATTTGGGCGGGTACGTCGCGCGGGAAGGCGGCTTACGCGCTTCGAGCTTGTTCGGA 563  
QY 753 GACGAAGAGATCACGCGCAACGCGCGAATGTTCAATGTCGTATCAACAAGCTTTCGCA 812  
Db 564 GACGAAGAGATCACGCGCAACGCGCGAATGTTCAATGTCGTATCAACAAGCTTTCGCA 623  
QY 813 CTCGACACTCCCTTACCAGACCTTCGAGTTTCGAGCCCGACGTAATCCGGTGTGTG 872  
Db 624 CTCGACACTCCCTTACCAGACCTTCGAGTTTCGAGCCCGACGTAATCCGGTGTGTG 683  
QY 873 GCGTGAGATCAAGACGCGACAAACACTTCCCTTCAAGCCGCGGCAAGTCAAGCCCATTTCA 932  
Db 684 GCGTGAGATCAAGACGCGACAAACACTTCCCTTCAAGCCGCGGCAAGTCAAGCCCATTTCA 743  
QY 933 CCGCGGACGATCACGGGGCTTGTAAAGCGCATGACGCTGACCGCGTGCACCGCGGGG 992  
Db 744 CCGCGGACGATCACGGGGCTTGTAAAGCGCATGACGCTGACCGCGTGCACCGCGGGG 803  
QY 993 CGAGACGATTGGGAAGAAGACCGCTTCAAGCGCGCTGGGACCGCGCAACCGTTATGCGAAT 1052  
Db 804 CGAGACGATTGGGAAGAAGACCGCTTCAAGCGCGCTGGGACCGCGCAACCGTTATGCGAAT 863  
QY 1053 CCTTCGGACCGCGGCTATTGCGGGCTTCGCGCTGAGGTGATCTTCAAGAAGAAGCCGGA 1112  
Db 864 CCTTCGGACCGCGGCTATTGCGGGCTTCGCGCTGAGGTGATCTTCAAGAAGAAGCCGGA 923  
QY 1113 CGGACGCGGACGACGAGATTTAGGGTTACCGCATTCAGCGCGGACCGCATCGCTCCG 1172  
Db 924 CGGACGCGGACGACGAGATTTAGGGTTACCGCATTCAGCGCGGACCGCATCGCTCCG 983  
QY 1173 GCCGGTCAAGCTTGATTGCGGACCGATCATGAGCCCGCTGAAGTGTATGAGCTTCAGGC 1232  
Db 984 GCCGGTCAAGCTTGATTGCGGACCGATCATGAGCCCGCTGAAGTGTATGAGCTTCAGGC 1043  
QY 1233 GTGGTTGACGGGCGAGGGGCGCGCAAGGGGCTTCCCGGGGGCAAGCCATTTGTCCGC 1292  
Db 1044 GTGGTTGACGGGCGAGGGGCGCGCAAGGGGCTTCCCGGGGGCAAGCCATTTGTCCGC 1103  
QY 1293 CATGACAAGCTGTACTGCGAGTGTGCGCGCTCATGACTTTCGAAGCGCGGGAAGAATC 1352  
Db 1104 CATGACAAGCTGTACTGCGAGTGTGCGCGCTCATGACTTTCGAAGCGCGGGAAGAATC 1163  
QY 1353 GATCAAGACTCTTACCGCTGCGCTGCGCGGAAGGTGTCGACCCGTCGCACTGGGCA 1412  
Db 1164 GATCAAGACTCTTACCGCTGCGCTGCGCGGAAGGTGTCGACCCGTCGCACTGGGCA 1223



OY	1353	GATCAAGGACTCTTACCCTGCCGTGC	CGGGAAGGTGGTCACCCGTC	CGCACCTGGGCA	1412
Db	1164	GATCAAGGACTCTTACCCTGCCGTGC	CGGGAAGGTGGTCACCCGTC	CGCACCTGGGCA	1223
OY	1413	GCACGAAGGCACGTGCACGTCAGCATGGCGGCAC	TGCACAAGTTGTTGCGAAGC	CAT	1472
Db	1224	GCACGAAGGCACGTGCACGTCAGCATGGCGGCAC	TGCACAAGTTGTTGCGAAGC	CAT	1283
OY	1473	CTTCAACAAGATCAGGCGACGCCGAAGCGACGAAGAC	GTGGCGCTTCTGTGGGAAC		1532
Db	1284	CTTCAACAAGATCAGGCGACGCCGAAGCGACGAAGAC	GTGGCGCTTCTGTGGGAAC		1343
OY	1533	CGCCCGACGCTTCGGGCAAGCTCACTGAGGCGCCTGAGA	AAGCGCGGCAACGGCGCAACT		1592
Db	1344	CGCCCGACGCTTCGGGCAAGCTCACTGAGGCGCCTGAGA	AAGCGCGGCAACGGCGCAACT		1403
OY	1593	TGTTGCGGAGCGCGCCGACGCCCCGTGAACGCCCTTGAA	GAGCTGTACGAAGACCGCGGCG		1652
Db	1404	TGTTGCGGAGCGCGCCGACGCCCCGTGAACGCCCTTGAA	GAGCTGTACGAAGACCGCGGCG		1463
.OY	1653	AGGCGCGTACGACGAGACCCGTTGGCAGGAAGCACTTCCG	GAACACGAGCAGCGCTGAC		1712
Db	1464	AGGCGCGTACGACGAGACCCGTTGGCAGGAAGCACTTCCG	GAACACGAGCAGCGCTGAC		1523
OY	1713	GCTCCGGCAGCAAGGGGCGGAAGAGCGGCTTGCCCACTTGA	AAGCCGCCGAAGCCCCGAA		1772
Db	1524	GCTCCGGCAGCAAGGGGCGGAAGAGCGGCTTGCCCACTTGA	AAGCCGCCGAAGCCCCGAA		1583
OY	1773	GCTTCCCCTTGACCAATGGTTCCCCGGAAGACGCCGACGCTGA	ACC CGGCGGCTTAAGTC		1832
Db	1584	GCTTCCCCTTGACCAATGGTTCCCCGGAAGACGCCGACGCTGA	ACC CGGCGGCTTAAGTC		1643
OY	1833	GTGGTGGGGGCGCGCTCAGTAGACGACAAGCGCGTGTCTGTGGG	CTTCTGTAGACAA		1892
Db	1644	GTGGTGGGGGCGCGCTCAGTAGACGACAAGCGCGTGTCTGTGGG	CTTCTGTAGACAA		1703
OY	1893	GATCGTTGTACGAAGTCGACTACGGGCGAGGGGCGAGGGAAC	CGCCCATCGAAGCGCGC		1952
Db	1704	GATTGTGTACGAAGTCGACTACGGGCGAGGGGCGAGGGAAC	CGCCCATCGAAGCGCGC		1763
OY	1953	TTGCAATCAGTGGGCGAAGCGCGCGACCGACGACGACGACA	GAGACGACGCGCCAGAGCGCAC		2012
Db	1764	TTGCAATCAGTGGGCGAAGCGCGCGACCGACGACGACGACA	GAGACGACGCGCCAGAGCGCAC		1822
OY	2013	GGAAGACGTAG	2023		
Db	1823	GGAAGACGTAG	1833		

```

RESULT 8
US-09-887-576-582
; Sequence 582, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 582
;
; LENGTH: 2000
;
; TYPE: DNA
;
; ORGANISM: Arabidopsis thaliana
;
US-09-887-576-582

```

```
Query Match      9.68; Score 195; DB 10; Length 2000;
Best Local Similarity 100.0%; Pred. NO. 2.5e-36;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	30	TACGTAAGTTCTCGCTTCACCTTGGATATATATATAAATTATCATTAATTAGTAGTA	89
Db	383	TACGTAAGTTCTCGCTTCACCTTGGATATATATAAATAATTATCATTAAATTAGTAGTA	442
QY	90	ATATAATATTTCAAATATTTTTTCCAAATAAAAGAATGTAGTATATAGCAATTGCCTTT	149
Db	443	ATATAATATTTCAAATATTTTTTCCAAATAAAAGAATGTAGTATATAGCAATTGCCTTT	502
QY	150	CTGTAGTTTATAAGTGGTATATTTTAAATTTATAACTTTCTAATATATGACCACCAAAATTT	209
Db	503	CTGTAGTTTATAAGTGGTATATTTTAAATTTATAACTTTCTAATATATGACCACCAAAATTT	562
QY	210	GTTGATGTGCAGGTA 224	
Db	563	GTTGATGTGCAGGTA 577	

RESULT 9  
US-09-88

; Sequence 581, Application US/09887576  
; Patent No. US20020144047A1

```

; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 581
; LENGTH: 11180
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-581

```

	Query Match	9.68;	Score 195;	DB 10;	Length 11180;	
	Best Local Similarity	100.0%;	Pred. No. 6.9e-36;			
	Matches 195;	Conservative	0;	Mismatches	0;	Indels
						Gaps 0;
QY	30 TACGTAAGTTTCGTCTTCTACCTTGATATATATAAATTATCATTAATTAGTAGTA					89
Db	2231 TACGTAAGTTTCGTCTTCTACCTTGATATATATAAATTATCATTAATTAGTAGTA					2290
QY	90 ATATAATATTTCAAATATTTTTTCCAAAAPAAAGAATGTAGTATATAGCAATTGCTTTT					149
Db	2291 ATATAATATTTCAAATATTTTTTCCAAAATAAAGAATGTAGTATATAGCAATTGCTTTT					2350
QY	150 CTGTAGTTTATAAGTGTATATATTTTAAATTATAACTTTTCTAATATATGACCAAAATTT					209
Db	2351 CTGTAGTTTATAAGTGTATATATTTTAAATTATAACTTTTCTAATATATGACCAAAATTT					2410

PRIOR APPLICATION NUMBER: 60/184,822

Best Local Similarity 57.88; Pred. No. 0.29;  
Matches 89; Conservative 0; Mismatches 65; Indels 0; Gaps 0.

QY      43 GCCTCACCCTTGATATAATATAATTAATTATCAAGTAAATATATTCACA 102  
         . | | | | | | | | | | | | | | | | | | | | |  
Db    11188 GTTTTGCGTTGTGTTTAGTGGATTTTTTTTTTTTATAGTTAAAAATGTATTTTTTAA 11247







```

;      2000-04-07
;      2000-06-30
;      2000-09-01
;      NUMBER OF SEQ ID NOS: 228
;      SEQ ID NO 117
;      LENGTH: 11036
;      TYPE: DNA
;      ORGANISM: Artificial Sequence
;      FEATURE:
;      OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;      FEATURE:
;      NAME/KEY: unsure
;      LOCATION: (1927..1928, 1943, 1956, 1958, 1981)
US-10-239-676-117

```

```

Query Match      2.48;   Score 49;   DB 9;   Length 11036;
Best Local Similarity 56.58;   Pred. No. 0.43;
Matches      91;   Conservative      0;   Mismatches      70;   Indels      0;   Gaps      0;

```

Qy	52	TTTGATATATATATAATTAATTAATCATTAATTAGTAGTAATATAATATATTCCAATATATTTT	111
Pb	8630	TTTGATATGAATTATTAGTTTTTTAATATTGGTTAAATTATAGATTTTATAGTATTTTA	8689
Qy	112	TTCAAAATAAAGAATGAGTATATAGCAATTCCTTCTGAGTTTATAAGTGCTATA	171
Pb	8690	TTAATAAAGAATAATATAAGTTAATTTATTAGTTTATATGCTATATAGTATATGTA	8749
Qy	172	TTTTAATTATAACCTTTCTAATATATGACCAAAATTTCTT	212
Pb	8750	TATTTATTTTAGTATTTAATGATTTTAATAAAATATTTT	8790

Search completed: May 11, 2003, 23:29:56  
Job time : 395.5 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 15:44:18 ; Search time 5977.5 Seconds

(without alignments)  
9888.381 Million cell updates/sec

Title: US-09-940-550A-10

Perfect score: 2031

Sequence: 1 atgcacacagggtgtgtac.....cggaagacgtagcgcgctag 2031

Scoring table: IDENTITY\_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2031	100.0	2031	6	AX394183 Sequence
2	2015	99.2	2031	6	AX394182 Sequence
3	1998	98.4	9901	6	AX394184 Sequence
4	1796.6	88.5	2200	7	C31INTDNA
5	1796.6	88.5	5711	6	AX114861 Sequence
6	1796.6	88.5	5715	12	PCL414670 Cloning v
7	1796.6	88.5	41489	7	BPH6589
8	200.4	9.9	5967	12	EVU84006
9	198.6	9.8	912	6	AX259239 Sequence
10	198.6	9.8	1577	6	AX259244 Sequence
11	198.6	9.8	1618	6	AX259242 Sequence
12	198.6	9.8	2943	6	AX259250 Sequence
13	195	9.6	2001	6	AX259240 Sequence
14	195	9.6	2001	6	AX356863 Sequence
15	195	9.6	2633	6	AR037156 Sequence
16	195	9.6	2633	6	AR112042 Sequence
17	195	9.6	2725	6	AX259245 Sequence
18	195	9.6	2730	6	AX259243 Sequence
19	195	9.6	4072	6	AX259251 Sequence
20	195	9.6	4341	6	AX259249 Sequence
21	195	9.6	5288	12	ASTDNABV
22	195	9.6	5534	6	AR037157 Sequence
23	195	9.6	5534	6	AR112043 Sequence
24	195	9.6	12817	6	AX356864 Sequence
25	195	9.6	13274	6	AX356862 Sequence
26	194	9.6	1701	6	AR174855 Sequence
27	194	9.6	3336	6	AR174856 Sequence
28	194	9.6	3877	6	AR174858 Sequence
29	190.2	9.4	5560	6	A60112 Sequence 5
30	190.2	9.4	5560	6	AR098311 Sequence
31	190	9.4	1034	6	A87679 Sequence 2
32	181.4	8.9	363	6	AX060392 Sequence
33	180.6	8.9	363	6	AX060389 Sequence
34	180.6	8.9	363	6	AX060391 Sequence
35	180.6	8.9	369	6	AX060390 Sequence
36	179.8	8.9	3699	8	STLSIG
37	62.2	3.1	39314	1	SGR300302
38	61.2	3.0	125020	9	AF429315 Homo sapi
39	60	3.0	125020	9	AF429315 Homo sapi
40	59	2.9	134375	2	AC118169 Rattus no
41	58.6	2.9	169546	2	AC004157 Plasmodi
42	58.2	2.9	154902	9	AC005799 Homo sapi
43	58.2	2.9	191947	2	AC079210 Homo sapi
44	58	2.9	9424	8	SD0404228 Saccharom
45	57.8	2.8	1181	8	MISCOX13 X14910 Yeast mitoc

ALIGNMENTS

RESULT 1	AX394183	2031 bp	DNA	linear	PAT 23-MAR-2002
LOCUS	AX394183	Sequence 10 from Patent WO0216609.			
DEFINITION	AX394183				
ACCESSION	AX394183				
VERSION	AX394183.1	GI:19702117			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE	1				
AUTHORS	Mankin, L.				
TITLE	Self-excising polynucleotides and uses thereof				
JOURNAL	Patent: WO 0216609-A 10 28-FEB-2002;				
	BASF Plant Science GmbH (DE) ; McKersie, Bryan (US)				

Filed at  
Instat US

FEATURES Location/Qualifiers  
source 1. 2031  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="C31int\* DNA sequence"  
BASE COUNT 454 a 558 c 640 g 379 t  
ORIGIN

Query Match 100.0%; Score 2031; DB 6; Length 2031;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2031; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCACAAAGGGGTGTGACCGGGGTGATACGATTTCTGCTTCTACCTTGATATA 60  
Db 1 ATGGCACAAAGGGGTGTGACCGGGGTGATACGATTTCTGCTTCTACCTTGATATA 60  
QY 61 TATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120  
Db 61 TATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120  
QY 121 AAGAATGTAGTATATAGCAATTTGCTTCTGATTTTATAGTGTATATTTTATTT 180  
Db 121 AAGAATGTAGTATATAGCAATTTGCTTCTGATTTTATAGTGTATATTTTATTT 180  
QY 181 ATAACTTTTCTAATATATGACCAAAATTTGTTGATGTGACAGTACCGGGTCTTAC 240  
Db 181 ATAACTTTTCTAATATATGACCAAAATTTGTTGATGTGACAGTACCGGGTCTTAC 240  
QY 241 CGTCAGTCGCGGAGCGGAGATAGCAGTGCAGCAAGCCCGACACACAGCGTAGCGCC 300  
Db 241 CGTCAGTCGCGGAGCGGAGATAGCAGTGCAGCAAGCCCGACACACAGCGTAGCGCC 300  
QY 301 AACGAAGACAAGGGCGGCGACCTTCAGCGCGAAGTGCAGCGCGGGGGCGGTTACAG 360  
Db 301 AACGAAGACAAGGGCGGCGACCTTCAGCGCGAAGTGCAGCGCGGGGGCGGTTACAG 360  
QY 361 TTGCTCGGCATTTACGGAAGCGCGCGCACGTCGGCGTTGCGGAGCGGGAGCGCGCG 420  
Db 361 TTGCTCGGCATTTACGGAAGCGCGCGCACGTCGGCGTTGCGGAGCGGGAGCGCGCG 420  
QY 421 GAGTTCGAACGCATCTTGAACGAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
Db 421 GAGTTCGAACGCATCTTGAACGAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
QY 481 GACGTGTCGCGCTTCTCGCGCTGAAGTGCATGACGCGCATTCGATTTGTCGGAATG 540  
Db 481 GACGTGTCGCGCTTCTCGCGCTGAAGTGCATGACGCGCATTCGATTTGTCGGAATG 540  
QY 541 CTCGCGCTGGGCGTGACGATTTGTTCCACTCAGGAAGCGCTTCCGCGAGGAAACGTC 600  
Db 541 CTCGCGCTGGGCGTGACGATTTGTTCCACTCAGGAAGCGCTTCCGCGAGGAAACGTC 600  
QY 601 ATGGACCTGATTCACCTGATTAATGCGGCTCGACGCGTCGCAAAAGATCTTCGTAAG 660  
Db 601 ATGGACCTGATTCACCTGATTAATGCGGCTCGACGCGTCGCAAAAGATCTTCGTAAG 660  
QY 661 TCGGCGAAGATTTCTGACACGAAGAACCTTCAGCGCGAATTTGGCGGTACGTCGCGGG 720  
Db 661 TCGGCGAAGATTTCTGACACGAAGAACCTTCAGCGCGAATTTGGCGGTACGTCGCGGG 720  
QY 721 AAGGCGCTTACGCGCTTCGAGCTTTTCGGAGACGAAGAGATCAGCGCGCAAGCGCCGA 780  
Db 721 AAGGCGCTTACGCGCTTCGAGCTTTTCGGAGACGAAGAGATCAGCGCGCAAGCGCCGA 780  
QY 781 ATGCTCAATGTGTCATCAACAAGTTAGCGCACTCGACCACTCCCTTACCGGAACCTTC 840  
Db 781 ATGCTCAATGTGTCATCAACAAGTTAGCGCACTCGACCACTCCCTTACCGGAACCTTC 840  
QY 841 GAGTTCGAGCCCGACGTAATCCGGTGTGTGCGGTGAGATCAAGACGCAACAAACACCTT 900  
Db 841 GAGTTCGAGCCCGACGTAATCCGGTGTGTGCGGTGAGATCAAGACGCAACAAACACCTT 900  
QY 901 CCCTTCAAGCCGGGAGTCAAGCCGCCATTCACCGCGGCGAGCATCACGGGGCTTTGTAG 960

Db 901 CCCTTCAAGCCGGGAGTCAAGCCGCCATTCACCGGGCGAGCATACGGGCTTTGTAG 960  
QY 961 CGCATGACGCTGACGCGGTGCCGACCCGGGCGAGACGATTTGGGAAGAACCGCTTCA 1020  
Db 961 CGCATGACGCTGACGCGGTGCCGACCCGCGAGACGATTTGGGAAGAACCGCTTCA 1020  
QY 1021 AGCGCTGGGACCCGGCAACCGTTATGCAATCTTCGGGACCCGCTATTGGCGGCTTC 1080  
Db 1021 AGCGCTGGGACCCGGCAACCGTTATGCAATCTTCGGGACCCGCTATTGGCGGCTTC 1080  
QY 1081 GCCGCTGAGTGATCTACAAGAAGACCGGACGCGACGCGGACCAAGATTTGAGGT 1140  
Db 1081 GCCGCTGAGTGATCTACAAGAAGACCGGAGACGCGGACCAAGATTTGAGGT 1140  
QY 1141 TACCGCATTCAGCGCGGACCCGATCAGCGCTCCGGCGGTGAGCTTGATTCGACCGCATC 1200  
Db 1141 TACCGCATTCAGCGCGGACCCGATCAGCGCTCCGGCGGTGAGCTTGATTCGACCGCATC 1200  
QY 1201 ATCGAGCCCGCTGAGTGTATGAGCTTCAGCGGTGTTGGACGGCAGGGGGCGCGCAAG 1260  
Db 1201 ATCGAGCCCGCTGAGTGTATGAGCTTCAGCGGTGTTGGACGGCAGGGGGCGCGCAAG 1260  
QY 1261 GGGCTTCCCGGGGCGCAAGCCATTTCTCCGCCATGAGCAAGCTGTACTGCGAGTGTGC 1320  
Db 1261 GGGCTTCCCGGGGCGCAAGCCATTTCTCCGCCATGAGCAAGCTGTACTGCGAGTGTGC 1320  
QY 1321 GCCGTATGACTTCGAAGCGCGGGGGAAGATCGATCAAGCACTTTACCGCTGCCGTGC 1380  
Db 1321 GCCGTATGACTTCGAAGCGCGGGGGAAGATCGATCAAGCACTTTACCGCTGCCGTGC 1380  
QY 1381 CGGAAGGTGCTGACCCCGTCCGCACTGGGCGAGCAGCAGGAAGCACTGCAACGTCAGCATG 1440  
Db 1381 CGGAAGGTGCTGACCCCGTCCGCACTGGGCGAGCAGCAGGAAGCACTGCAACGTCAGCATG 1440  
QY 1441 GCGGCACTGCACAAGTTCGTTGCGGAACGCATCTTCAACAAGATCAGGACGCGGAAGGC 1500  
Db 1441 GCGGCACTGCACAAGTTCGTTGCGGAACGCATCTTCAACAAGATCAGGACGCGGAAGGC 1500  
QY 1501 GACGAAGACGCTTGCGCTTGTGGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560  
Db 1501 GACGAAGACGCTTGCGCTTGTGGGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560  
QY 1561 GCGCTGAGAAGAGCGGCGAAGCGGCGAACCCTTGTGCGGAAGCGCGCGCGCGCGCTGAAC 1620  
Db 1561 GCGCTGAGAAGAGCGGCGAAGCGGCGAACCCTTGTGCGGAAGCGCGCGCGCGCGCTGAAC 1620  
QY 1621 GCCCTTGAAGAGCTGTACGAAGACCGCGCGGACGAGCTTACGACGACCCGTTGGCAGG 1680  
Db 1621 GCCCTTGAAGAGCTGTACGAAGACCGCGCGGACGAGCTTACGACGACCCGTTGGCAGG 1680  
QY 1681 AAGCACTTCGGGAAGCAACAGCGCGCTGACGCTCCGGCAGCAAGGGGCGGAAGAGCGG 1740  
Db 1681 AAGCACTTCGGGAAGCAACAGCGCGCTGACGCTCCGGCAGCAAGGGGCGGAAGAGCGG 1740  
QY 1741 CTGCGCAACTTGAAGCCCGCAAGCCCGGAAGTTGCCCTTGACCAATGCTTCCCGAA 1800  
Db 1741 CTGCGCAACTTGAAGCCCGCAAGCCCGGAAGTTGCCCTTGACCAATGCTTCCCGAA 1800  
QY 1801 GAGCGGACGCTGACCCGACCGGCGCTTAAGTGTGTGGGGCGCGCGCTGAGACGAC 1860  
Db 1801 GAGCGGACGCTGACCCGACCGGCGCTTAAGTGTGTGGGGCGCGCGCTGAGACGAC 1860  
QY 1861 AAGCGGCTGTTGTCGCGGCTCTTGTAGACAGATCGTTGTACAGAACTGACTACGGGC 1920  
Db 1861 AAGCGGCTGTTGTCGCGGCTCTTGTAGACAGATCGTTGTACAGAACTGACTACGGGC 1920  
QY 1921 AAGGGGCGAGGGAACGCCATCGAAGAGCGCGCTTCGATCAGTGGGCGAAGCCGCGAC 1980  
Db 1921 AAGGGGCGAGGGAACGCCATCGAAGAGCGCGCTTCGATCAGTGGGCGAAGCCGCGAC 1980  
QY 1981 GACGACGACGAAGACGACCGCGAGGACGCGAAGAGAGCTAGCGGCTAG 2031





QY 1741 CTTGCCGAAGTGAAGCCCGGAGCCCGGAGCTTCCCTTGACCAATGTTCCCGGAA 1800  
|||||  
Db 1741 CTTGCCGAAGTGAAGCCCGGAGCCCGGAGCTTCCCTTGACCAATGTTCCCGGAA 1800  
QY 1801 GACGCCGAGCTGACCCGACCGCCCTAAGTCTGTGGGGCGCGCTCAGTAGACGAC 1860  
|||||  
Db 1801 GACGCCGAGCTGACCCGACCGCCCTAAGTCTGTGGGGCGCGCTCAGTAGACGAC 1860  
QY 1861 AAGCGCGTGTCTGCTGGGCTCTCTGAGACAGATCGTTGTACGAGTCTGACTACGGGC 1920  
|||||  
Db 1861 AAGCGCGTGTCTGCTGGGCTCTCTGAGACAGATCGTTGTACGAGTCTGACTACGGGC 1920  
QY 1921 AGGGGGCAGGGAACGCCCATCGAAGCGCGCTTCATCAGTGGGGGGAAGCCCGGAC 1980  
|||||  
Db 1921 AGGGGGCAGGGAACGCCCATCGAAGCGCGCTTCATCAGTGGGGGGAAGCCCGGAC 1980  
QY 1981 GACGACGACGAAGACGACCGCCAGGACGGCAGGAAGACGTAGCGCGCTAG 2031  
|||||  
Db 1981 GACGACGACGAAGACGACCGCCAGGACGGCAGGAAGACGTAGCGCGCTAG 2031

RESULT 3  
AX394184  
LOCUS AX394184 9901 bp DNA linear PAT 23-MAR-2002  
DEFINITION Sequence 11 from Patent WO0216609.  
ACCESSION AX394184  
VERSION AX394184.1 GI:19702118  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct  
REFERENCE 1  
AUTHORS Mankin, L.  
TITLE Self-excising polynucleotides and uses thereof  
JOURNAL Patent: WO 0216609-A 11 28-FEB-2002;  
FEATURES  
source location/Qualifiers  
1. 9901  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="PBPS EW051 T-DNA region"  
BASE COUNT 2489 a 2371 c 2478 g 2551 t. 12 others  
ORIGIN

Query Match 98.4%; Score 1998; DB 6; Length 9901;  
Best local similarity 99.2%; Pred. No. 0;  
Matches 2021; Conservative 0; Mismatches 10; Indels 7; Gaps 1;

QY 1 ATGGCACAAGGGGTTGTGACCGGGGTGATAGTAAGTTTCTGCTTCTACCTTTGATATA 60  
|||||  
Db 4984 ATGGCACAAGGGGTTGTGACCGGGGTGATAGTAAGTTTCTGCTTCTACCTTTGATATA 5043  
QY 61 TATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120  
|||||  
Db 5044 TATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5103  
QY 121 AAAGAATGTAGTATATAGCAATGCTTTCTGTAGTTTAAGTGTATATTTTAATTT 180  
|||||  
Db 5104 AAAGAATGTAGTATATAGCAATGCTTTCTGTAGTTTAAGTGTATATTTTAATTT 5163  
QY 181 ATAATCTTTCTATATATATGACCAAAATTTGTTGATGTGACGAGCGGGTCTTACGAC 240  
|||||  
Db 5164 ATAATCTTTCTATATATATGACCAAAATTTGTTGATGTGACGAGCGGGTCTTACGAC 5223  
QY 241 CGTC-----AGTCGGCGAGCGCGAGATAGCAGTGCAGCAAGCCGACACAGCG 293  
|||||  
Db 5224 CGTCGRCNTNDAGTCGCGGAGCGCGAGAAATTCAGCGCAGCAAGCCGACACAGCG 5283  
QY 294 TAGCGCCAAGAGACAGAGCGCGCGACCTTCAGCGCGGAAGTGCAGCGCGACGGGGCGG 353  
|||||  
Db 5284 TAGCGCCAAGAGACAGAGCGCGCGCGACCTTCAGCGCGGAAGTGCAGCGCGACGGGGCGG 5343  
QY 354 GTTCAGGTTCTGTCGGGCAATTTACGGAAGCGCGCGGACGTCGGCGTTCCGGACGGCGGA 413

Db 5344 GTTCAGGTTCTGTCGGGCAATTTACGGAAGCGCGCGGACGTCGGCGTTCCGGACGGCGGA 5403  
QY 414 GCGCCCGAGTTGCAACGCATCTGAACGAATGCCGCGCGCGCGCTCAACATGATCAT 473  
|||||  
Db 5404 GCGCCCGAGTTGCAACGCATCTGAACGAATGCCGCGCGCGCGCTCAACATGATCAT 5463  
QY 474 TGTCTATGACGTGTGCGGCTTCTGCGGCTGAAGGTATGAGCGGATTCGATTTCTC 533  
|||||  
Db 5464 TGTCTATGACGTGTGCGGCTTCTGCGGCTGAAGGTATGAGCGGATTCGATTTCTC 5523  
QY 534 GGAATTTGCTGCGGCTTGGGCGTGACGATTTGTTCCACTCAGGAAGCGCTTCCGACGAG 593  
|||||  
Db 5524 GGAATTTGCTGCGGCTTGGGCGTGACGATTTGTTCCACTCAGGAAGCGCTTCCGACGAG 5583  
QY 594 AAACGTATGACCTGATTCACCTGATTTATGCGGCTCGACCGCTCGCACAAAGATCTTC 653  
|||||  
Db 5584 AAACGTATGACCTGATTCACCTGATTTATGCGGCTCGACCGCTCGCACAAAGATCTTC 5643  
QY 654 GCTGAGTCGCGGAAGATTTCTCGACACGAAGAACCCTTCAGCGCGGAATTTGGCGGTAAGT 713  
|||||  
Db 5644 GCTGAGTCGCGGAAGATTTCTCGACACGAAGAACCCTTCAGCGCGGAATTTGGCGGTAAGT 5703  
QY 714 CGCGGGGAAGCGGCTTACGGCTTCGAGCTTGTTCGAGAGCAAGAGATACGCGCAA 773  
|||||  
Db 5704 CGCGGGGAAGCGGCTTACGGCTTCGAGCTTGTTCGAGAGCAAGAGATACGCGCAA 5763  
QY 774 CGGCGGAATGTCATATGTCGATCAACAAGTTAGCGCACTGCACCACTCCCTTACCGG 833  
|||||  
Db 5764 CGGCGGAATGTCATATGTCGATCAACAAGTTAGCGCACTGCACCACTCCCTTACCGG 5823  
QY 834 ACCCTTCGAGTTCGAGCCCGGACGTAATCCGGTGTGTGGCTGAGATCAAGACGCACAA 893  
|||||  
Db 5824 ACCCTTCGAGTTCGAGCCCGGACGTAATCCGGTGTGTGGCTGAGATCAAGACGCACAA 5883  
QY 894 ACACCTTCCCTTCAAGCCGGCGAGTCAAGCCGCAATTCACCCGGGACAGATCACGGGCT 953  
|||||  
Db 5884 ACACCTTCCCTTCAAGCCGGCGAGTCAAGCCGCAATTCACCCGGGACAGATCACGGGCT 5943  
QY 954 TTGTAAGCGCATGAGCGCTGACCGCGGTGCGGACCGGGGGGAGACGATTTGGGAAGAAGAC 1013  
|||||  
Db 5944 TTGTAAGCGCATGAGCGCTGACCGCGGTGCGGACCGGGGGGAGACGATTTGGGAAGAAGAC 6003  
QY 1014 CGCTTCAAGCGGCTTGGACCGCGGCAACCGTTATGCGAATCTCTCGGACCGCGCTATTGC 1073  
|||||  
Db 6004 CGCTTCAAGCGGCTTGGACCGCGGCAACCGTTATGCGAATCTCTCGGACCGCGCTATTGC 6063  
QY 1074 GGGCTTCGCGGCTGAGGTGATCTACAAGAAGAAGCCGGAACGCGGACCGGCAACGAAGAT 1133  
|||||  
Db 6064 GGGCTTCGCGGCTGAGGTGATCTACAAGAAGAAGCCGGAACGCGGACCGGCAACGAAGAT 6123  
QY 1134 TGAGGTTACCGCATTCAGCGCGGACCGGATCAGGCTCCGGCGGTGAGCTTGATTGGCG 1193  
|||||  
Db 6124 TGAGGTTACCGCATTCAGCGCGGACCGGATCAGGCTCCGGCGGTGAGCTTGATTGGCG 6183  
QY 1194 ACCGATCATGAGCGCGGCTGAGTGTATGAGCTTCAGGCGTGTGGACGGCGAGGGGCG 1253  
|||||  
Db 6184 ACCGATCATGAGCGCGGCTGAGTGTATGAGCTTCAGGCGTGTGGACGGCGAGGGGCG 6243  
QY 1254 CGGCAAGGGGCTTCCCGGGGGGCAAGCCATTTCTCCGCCATGGAACAAGCTGTACTGCGA 1313  
|||||  
Db 6244 CGGCAAGGGGCTTCCCGGGGGGCAAGCCATTTCTCCGCCATGGAACAAGCTGTACTGCGA 6303  
QY 1314 GTGTGGCGCGGCTGATGACTTCGAAGCGCGGGGAAGATGATCAAGGACTCTTACCGCTG 1373  
|||||  
Db 6304 GTGTGGCGCGGCTGATGACTTCGAAGCGCGGGGAAGATGATCAAGGACTCTTACCGCTG 6363  
QY 1374 CCGTCGCGGGAAGGTGTCGACCCGTCGCACTGGGACAGCAGCAAGGACAGTGAACGT 1433  
|||||  
Db 6364 CCGTCGCGGGAAGGTGTCGACCCGTCGCACTGGGACAGCAGCAAGGACAGTGAACGT 6423  
QY 1434 CAGCATGGCGGCACTGACAAGTTCGTTGCGGAAGCGATCTTCAACAAGATCAGGCACGC 1493  
|||||

Db 6424 CAGCATGGCGGCGACTCGACCAAGTTCGTTGCGGAACGCATCTTCAACAAGATCAAGGCACGC 6483  
Qy 1494 CGAAGCGCAGCAAGAGACGTTGGCGCTTCTGTGGAAAGCCGCCGAGCGCTTCGGCAAGCT 1553  
Db 6484 CGAAGCGCAGCAAGAGACGTTGGCGCTTCTGTGGAAAGCCGCCGAGCGCTTCGGCAAGCT 6543  
Qy 1554 CACTGAGCGCGCTGAGAAAGCGCGCAAGCGGCGCAAGCTTGTGGAGAGCGCGCGCAAGCGC 1613  
Db 6544 CACTGAGCGCGCTGAGAAAGCGCGCAAGCGGCGCAAGCTTGTGGAGAGCGCGCGCAAGCGC 6603  
Qy 1614 CCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGCAAGAGCTTACGACGCAACCGCT 1673  
Db 6604 CCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGCAAGAGCTTACGACGCAACCGCT 6663  
Qy 1674 TGGCAGGAAGCACTCCGGAAGCAACAGGCGCGCTGACGCTCCGCGAGCAAGCGCGCA 1733  
Db 6664 TGGCAGGAAGCACTCCGGAAGCAACAGGCGCGCTGACGCTCCGCGAGCAAGCGCGCA 6723  
Qy 1734 AGAGCGCGCTTGGCAACTTGAAGCGCGCGCAAGCGCGCAAGCTTGCCTTACCAATGCTT 1793  
Db 6724 AGAGCGCGCTTGGCAACTTGAAGCGCGCGCAAGCGCGCAAGCTTGCCTTACCAATGCTT 6783  
Qy 1794 CCCCAGAGACGCGCAGCGCTGACCCGACCGCGCTTAAGTCTGCTGGGGCGCGCTCAGT 1853  
Db 6784 CCCCAGAGACGCGCAGCGCTGACCCGACCGCGCTTAAGTCTGCTGGGGCGCGCTCAGT 6843  
Qy 1854 AGACGACAAGCGCGTGTCTGCTGGCGCTCTTCTGACACAGATCGTTGTACCAAGTCGAC 1913  
Db 6844 AGACGACAAGCGCGTGTCTGCTGGCGCTCTTCTGACACAGATCGTTGTACCAAGTCGAC 6903  
Qy 1914 TACGGGCAAGGGGGGAGGAGCGCGCTTGCATCAGCTGCGGCAAGCC 1973  
Db 6904 TACGGGCAAGGGGGGAGGAGCGCGCTTGCATCAGCTGCGGCAAGCC 6963  
Qy 1974 GCCGACGACGACGACGAGCAAGACGCGCCAGGAGCGGCAAGAGAGCTAGCGCGCTAG 2031  
Db 6964 GCCGACGACGACGACGAGCAAGACGCGCCAGGAGCGGCAAGAGAGCTAGCGCGCTAG 7021

RESULT 4  
C31INTDNA

LOCUS C31INTDNA 2200 bp DNA linear PHG 11-FEB-1992  
DEFINITION Bacteriophage phiC31 ORF 613 for p68.  
ACCESSION X59938  
VERSION X59938.1 GI:14903  
KEYWORDS integrase; orf 613; p68 gene; site-specific recombination.  
SOURCE Bacteriophage phi-C31.  
ORGANISM Bacteriophage phi-C31  
Virus; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
Lambda-like viruses.

REFERENCE 1 (bases 1 to 2200)  
AUTHORS Kuhstoss, S.A.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUN-1991) S.A. Kuhstoss, Lilly Research Labs, Lilly  
Corporate Center, Indianapolis, IN 46285, USA  
REFERENCE 2 (bases 1 to 2200)  
AUTHORS Kuhstoss, S. and Rao, R.N.  
TITLE Analysis of the integration function of the streptomyces  
bacteriophage phi C31

JOURNAL J. Mol. Biol. 222 (4), 897-908 (1991)  
MEDLINE 92106348  
PUBMED 1762157

FEATURES  
source

1. .2200 Location/Qualifiers  
/organism="Bacteriophage phi-C31"  
/db\_xref="taxon:10719"  
/map="38168-40368nt (standard phiC31 map)"  
/clone="pKC912"  
99. .137  
/note="putative terminator"  
/rpt\_type=INVERTED  
142. .168  
repeat\_region  
/rpt\_type=DIRECT

promoter 163. .190  
repeat\_region 178. .215  
misc\_recomb /note="putative integrase binding site"  
/rpt\_type=INVERTED  
196. .198  
prim\_transcript (202.203). >2073  
RBS 218. .222  
gene 232. .2073  
CDS 232. .2073  
/gene="ORF613"  
232. .2073  
/function="Integrase, or other factor essential for  
integration"  
/codon\_start=1  
/transl\_table=11  
/product="p68"  
/protein\_id="CAA42561.1"  
/db\_xref="GI:14904"  
/db\_xref="SPTREMBL:Q38194"  
/translation="MTQGVTVGVDYAGAYDROSERENSSAASAPATORSANEDKAAD  
LQREVERDGRFVGHSEAPGTSFETAEERPERILNECRAGRLNMIIVDVSRF  
SLKVMDAIPRIVSELALGVITVSTQEGVFRGNVMDLILHMRDASHKESLSAK  
ILDTKNLORELGVGKAPYGFELVSEETLRNGRMVNVINKLAHSTPLTGFPE  
FEPDIVRMMWREIKTHKHLPEKPSQAAIHPSITGLCKRMDADAVPTGETIGKTA  
SSAMDPAVMRILRDPRIAGFAEVIYKKRDPPTTKIEGYRIODPITLRPELDC  
GPRIEPAWVELQAWLDGRGRKGLSRQALISAMDLYCECAVMTSKRGEESIKDS  
YRCRRKRVDPSPAPGHEGTNCVSMALDKFAERIFNKIRHAEGETLALWENAR  
RFGKLTAEPEKSGERANLVAERADALNLEELYEDRAGAYDGPVGRKHKRQQAALT  
LROGAEERLAELEAPKPLIDQWPEADADPTGPKSMWGRASVDKRVFVGLFV  
DKIVTKSTGRGGGTPIEKRASITWAKPTDDEDDAQQDGEDVAA"

repeat\_region

BASE COUNT 441 a 659 c 738 g 362 t  
ORIGIN

Query Match 88.5%; Score 1796.6; DB 7; Length 2200;  
Best local similarity 99.2%; Pred. No. 0;  
Matches 1805; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 213 GATGTGACAGTACGCGGGTGTACGACCGTACGTGCGCGCGAGCGGAGATAGACGTGC 272  
Db 255 GGTGACACGTACGCGGGTGTACGACCGTACGTGCGCGCGAGCGGAGATTCAGCGC 314  
Qy 273 AGCAAGCCGACGACACAGCGTAGCGCAAGCAAGAGCGCGCGACCTTCAGCGCA 332  
Db 315 AGCAAGCCGACGACACAGCGTAGCGCAAGCAAGAGCGCGCGACCTTCAGCGCA 374  
Qy 333 AGTCAGCGGACGCGGGCGGTTCAAGTTCTGCGGCGATTCAGCGAGCGCGGCGAC 392  
Db 375 AGTCAGCGGACGCGGGCGGTTCAAGTTCTGCGGCGATTCAGCGAGCGCGGCGAC 434  
Qy 393 GTCGGCGTTGCGGACGCGCGGAGCGCGGAGTTGCAACGCATCTGAACGAATGCCGCGC 452  
Db 435 GTCGGCGTTGCGGACGCGCGGAGCGCGGAGTTGCAACGCATCTGAACGAATGCCGCGC 494  
Qy 453 CGGGCGGCTCAACATGATGTTCTATGACGTTGCGCGCTTCTGCGCGCTGAAGGTCAT 512  
Db 495 CGGGCGGCTCAACATGATGTTCTATGACGTTGCGCGCTTCTGCGCGCTGAAGGTCAT 554  
Qy 513 GGACGCGATTCGATGTTCTCGGAATGCTCGCCCTGGCGCGTGAAGATTTTCCACTCA 572  
Db 555 GGACGCGATTCGATGTTCTCGGAATGCTCGCCCTGGCGCGTGAAGATTTTCCACTCA 614  
Qy 573 GGAAGCGTTCGCGGACGGAAGACGTGACCTGATTCACCTGATATGCGGCTCGA 632  
Db 615 GGAAGCGTTCGCGGACGGAAGACGTGACCTGATTCACCTGATATGCGGCTCGA 674  
Qy 633 CGCGTCGACAAGATCTTCGCTGAAGTCGGCGGAGATTCGACACGAGAAGACCTTCA 692  
Db 675 CGCGTCGACAAGATCTTCGCTGAAGTCGGCGGAGATTCGACACGAGAAGACCTTCA 734

```
QY 693 GCCGAATTGGCGGGTACGTCCGGCGGAAGGCCCTTACGGCTTCGAGCTTGTTCGGA 752
Db 735 GCGCGAATTGGCGGGTACGTCCGGCGGAAGGCCCTTACGGCTTCGAGCTTGTTCGGA 794
QY 753 GACGAAGAGATCACGCGCAACGGCCGAATGTCATGTCTCATCAACAAGTTAGCCGA 812
Db 795 GACGAAGAGATCACGCGCAACGGCCGAATGTCATGTCTCATCAACAAGCTTGCCTGA 854
QY 813 CTCGACCCTCTCCCTTACCCGACCCCTTCGAGTTCCAGCCCGACGTATCCGGTGTGTG 872
Db 855 CTCGACCCTCTCCCTTACCCGACCCCTTCGAGTTCCAGCCCGACGTATCCGGTGTGTG 914
QY 873 GCGTGAGATCAAGACGCAACAACACCTTCCCTTCAAGCCGGGACGTCAAGCCGCTTCA 932
Db 915 GCGTGAGATCAAGACGCAACAACACCTTCCCTTCAAGCCGGGACGTCAAGCCGCTTCA 974
QY 933 CCGGGGACGATCACGGGGCTTTGTAAGCGCATGGACGCTGACGCGCTGCCGACCCGGGG 992
Db 975 CCGGGGACGATCACGGGGCTTTGTAAGCGCATGGACGCTGACGCGCTGCCGACCCGGGG 1034
QY 993 CGAGACGATTGGGAAGAACCGCTTCAAGCCCTGGACCCCGGCAACCGTTATGCGAAT 1052
Db 1035 CGAGACGATTGGGAAGAACCGCTTCAAGCCCTGGACCCCGGCAACCGTTATGCGAAT 1094
QY 1053 CCTTCGGGACCCGCGTATTCGGGCTTCGCGCTGAGGTGATCTACAAGAAGACCGGA 1112
Db 1095 CCTTCGGGACCCGCGTATTCGGGCTTCGCGCTGAGGTGATCTACAAGAAGACCGGA 1154
QY 1113 CGGACGCGCCGACCAAGATTGAGGGTTACCGCATTCAGCGCGACCCGATCAGCGTCCG 1172
Db 1155 CGGACGCGCCGACCAAGATTGAGGGTTACCGCATTCAGCGCGACCCGATCAGCGTCCG 1214
QY 1173 GCGGCTCGAGCTTGAATTCGGGACCGATCATCGAGCCCGCTGAGTGTATGAGCTTCAGG 1232
Db 1215 GCGGCTCGAGCTTGAATTCGGGACCGATCATCGAGCCCGCTGAGTGTATGAGCTTCAGG 1274
QY 1233 GTGGTTGACGCGGAGGGGGCGCGGCAAGGGGCTTCCCGGGGCAAGCCATTCTGTCCG 1292
Db 1275 GTGGTTGACGCGGAGGGGGCGCGGCAAGGGGCTTCCCGGGGCAAGCCATTCTGTCCG 1334
QY 1293 CATGACAAGCTGTACTGCGAGTGTGGCGCCGTCACTGACTTCGAAGCGCGGGAAGATC 1352
Db 1335 CATGACAAGCTGTACTGCGAGTGTGGCGCCGTCACTGACTTCGAAGCGCGGGAAGATC 1394
QY 1353 GATCAAGGACTTTTACCGCTGCGCGCGGAAGGTGTGACCCGCTCCGACCTGGGCA 1412
Db 1395 GATCAAGGACTTTTACCGCTGCGCGCGGAAGGTGTGACCCGCTCCGACCTGGGCA 1454
QY 1413 GCACGAAGGACGTGCAACGTGACGATGGCGGCACTCGACAAGTTGCTGCGGAAGCAT 1472
Db 1455 GCACGAAGGACGTGCAACGTGACGATGGCGGCACTCGACAAGTTGCTGCGGAAGCAT 1514
QY 1473 CTTCAACAAGATCAGGACGCGGAAGGCGAAGAGAGACGTTGGCGCTTCTGTGGGAAGC 1532
Db 1515 CTTCAACAAGATCAGGACGCGGAAGGCGAAGAGAGACGTTGGCGCTTCTGTGGGAAGC 1574
QY 1533 CGCCCGACGCTTCGGCAAGCTCACTGAGGGCGCTGAGAGAAGCGGGAACCGGCAACT 1592
Db 1575 CGCCCGACGCTTCGGCAAGCTCACTGAGGGCGCTGAGAGAAGCGGGAACCGGCAACT 1634
QY 1593 TGTTCGGAGCGCGCGGACGCGCTTGAAGCGCTTGAAGAGCTGTACGAAGACGCGCGGC 1652
Db 1635 TGTTCGGAGCGCGCGGACGCGCTTGAAGCGCTTGAAGAGCTGTACGAAGACGCGCGGC 1694
QY 1653 AGGAGCTTACGACGACCCGTTGCGAAGACACTTCCGGAAGCAACAGGACGCGCTGAC 1712
Db 1695 AGGAGCTTACGACGACCCGTTGCGAAGACACTTCCGGAAGCAACAGGACGCGCTGAC 1754
QY 1713 GCTCCGCGACGAAGGGCGGAGAGCGGCTTCCGCACTTGAAGCGCGCGGCAAGCCCGAA 1772
Db 1755 GCTCCGCGACGAAGGGCGGAGAGCGGCTTCCGCACTTGAAGCGCGCGGCAAGCCCGAA 1814
QY 1773 GTTGCCTTTGACCAATGTTCCCGGAAGAGCGCGCTGACCCGACCGCGCTTAAGTC 1832
```

```
Db 1815 GCTTCCCTTGACCAATGGTTCCCGGAGACGCCGACGCTGACCCGACCGCCCTAAGTC 1874
QY 1833 GTGGTGGGGCGCGGCTCAGTAGACGACAAAGCGGCTGTTCGTCGGGCTCTCGTAGACAA 1892
Db 1875 GTGGTGGGGCGCGGCTCAGTAGACGACAAAGCGGCTGTTCGTCGGGCTCTTCCTAGACAA 1934
QY 1893 GATCGTTGTACGAAGTGCAGTACGGGCGAGGGGCGAGGGAAGCCCATCGAGAAGCGGC 1952
Db 1935 GATCGTTGTACGAAGTGCAGTACGGGCGAGGGGCGAGGGAAGCCCATCGAGAAGCGGC 1994
QY 1953 TTCGATCAGCTGGGCGGAAGCCCGGACGACGACGACGAAAGACGACGACGACGACGAC 2012
Db 1995 TTCGATCAGCTGGGCGGAAGCCCGGACGACGACGACGAAAGACGACGACGACGACGAC 2054
QY 2013 GGAAGACGTAGCGCGTAG 2031
Db 2055 GGAAGACGTAGCGCGTAG 2073
```

```
RESULT 5
AX114861 5711 bp DNA linear PAT 11-MAY-2001
LOCUS AX114861
```

```
DEFINITION Sequence 21 from Patent WO0129208.
ACCESSION AX114861
VERSION AX114861.1 GI:14031803
KEYWORDS
```

```
SOURCE
ORGANISM synthetic construct.
artificial construct.
1 (bases 1 to 5711)
```

```
REFERENCE
AUTHORS Kuehn, R., von Melchner, H. and Altschmied, J.
TITLE Conditional gene trapping construct for the disruption of genes
JOURNAL Patent: WO 0129208-A 21 26-APR-2001;
ARTEMIS Pharmaceuticals GmbH (DE); Frankgen Biotechnologie AG (DE)
```

```
FEATURES
source
1..5711
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="vector PRK65"
```

```
BASE COUNT 1329 a 1511 c 1617 g 1254 t
ORIGIN
```

```
Query Match 88.5%; Score 1796.6; DB 6; Length 5711;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1805; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

```
QY 213 GATGTGACGTACGCGGCTGCTTACGACCGTCACTCGCGGCGAGCGGAGAAATGACAGTGC 272
Db 1001 GGTGACACGTACGCGGCTGCTTACGACCGTCACTCGCGGCGAGCGGAGAAATGACAGTGC 1060
QY 273 AGCAAGCCGACGACAGCGGTAGCGCCCAACGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332
Db 1061 AGCAAGCCGACGACAGCGGTAGCGCCCAACGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1120
QY 333 AGTCGACGCGGACGCGGCGGCTTCAAGTTCTGCGGCAATTTCAGCGAAGCGCGGCGAC 392
Db 1121 AGTCGACGCGGACGCGGCGGCTTCAAGTTCTGCGGCAATTTCAGCGAAGCGCGGCGAC 1180
QY 393 GTCGCGCTTCGGGACGCGGAGCGCGCGGAGTTGCAACGCATCTGAACGAATGCCGCGC 452
Db 1181 GTCGCGCTTCGGGACGCGGAGCGCGCGGAGTTGCAACGCATCTGAACGAATGCCGCGC 1240
QY 453 CGGCGGCTCAACATGATGATTTGTATGACGTTGCGGCTTCTCGCGCTGAAGTCAAT 512
Db 1241 CGGCGGCTCAACATGATGATTTGTATGACGTTGCGGCTTCTCGCGCTGAAGTCAAT 1300
QY 513 GAGCGGATTCGATTTGTCGGAATTCGCGCCCTGGGCGTGACGATGTTTCCACTCA 572
Db 1301 GAGCGGATTCGATTTGTCGGAATTCGCGCCCTGGGCGTGACGATGTTTCCACTCA 1360
QY 573 GGAAGCGCTTTCGCGGAGGGAACGTCACTGACCTGATTCACCTGATTAAGCGGCTGA 632
```



Db 1361 GGAAGCGCTTCCGGCAGGAAACGTATGACCTGATTCACCTGATATATGCGGCTCGA 1420  
QY 633 CGCGTCGACAAAGAATCTTGGCTGAAGTCGGCGAAGATTCTGACACGAAGAACCCTTCA 692  
Db 1421 CGCGTCGACAAAGAATCTTGGCTGAAGTCGGCGAAGATTCTGACACGAAGAACCCTTCA 1480  
QY 693 GCGCGAATTGGGGGGGTACGTCGGCGGGGAAGGGCGCTTACGGCTTCGAGCTTGTTCGGA 752  
Db 1481 GCGCGAATTGGGGGGGTACGTCGGCGGGGAAGGGCGCTTACGGCTTCGAGCTTGTTCGGA 1540  
QY 753 GACGAAGAGATCACGCGCAACGGCCGAATGTGTCATATGTCATCAACAAGTTAGCGCA 812  
Db 1541 GACGAAGAGATCACGCGCAACGGCCGAATGTGTCATATGTCATCAACAAGTTAGCGCA 1600  
QY 813 CTGACCACTCCCTTACCGGACCTTCGAGTTCGAGCCGACGTAATCCGGTGTGTCG 872  
Db 1601 CTGACCACTCCCTTACCGGACCTTCGAGTTCGAGCCGACGTAATCCGGTGTGTCG 1660  
QY 873 GCGTGAGATCAAGACGACAAACACCTTCCCTTCAAGCCGGGAGTCAAGCCGCCATTCA 932  
Db 1661 GCGTGAGATCAAGACGACAAACACCTTCCCTTCAAGCCGGGAGTCAAGCCGCCATTCA 1720  
QY 933 CCGGGGACGATCACGGGGCTTGTAAAGCGCATGAGCGCTGACGCCGTGCCGACCCGGGG 992  
Db 1721 CCGGGGACGATCACGGGGCTTGTAAAGCGCATGAGCGCTGACGCCGTGCCGACCCGGGG 1780  
QY 993 CGAGACGATTGGGAAGAAGACCCCTTCAAGCGCCCTGGGACCCCGGCAACCCTTATGCGAAT 1052  
Db 1781 CGAGACGATTGGGAAGAAGACCCCTTCAAGCGCCCTGGGACCCCGGCAACCCTTATGCGAAT 1840  
QY 1053 CCTTCGGGACCCCGGTATTCGGGCTTCGCCGTGAGTGATCTACAAGAAGAGCCGGA 1112  
Db 1841 CCTTCGGGACCCCGGTATTCGGGCTTCGCCGTGAGTGATCTACAAGAAGAGCCGGA 1900  
QY 1113 CGGACGCGGACCCAGGAAGATTAGGGTTACCGCATTCAGCGCGGACCCGATCAGCTCCG 1172  
Db 1901 CGGACGCGGACCCAGGAAGATTAGGGTTACCGCATTCAGCGCGGACCCGATCAGCTCCG 1960  
QY 1173 GCCGTCGAGCTTGATTGGGACCGATCATCGAGCCCGCTGAGTGTATGAGCTTCAGGC 1232  
Db 1961 GCCGTCGAGCTTGATTGGGACCGATCATCGAGCCCGCTGAGTGTATGAGCTTCAGGC 2020  
QY 1233 GTGTTGACGGGACAGGGGGCGGGCAAGGGGCTTCCCGGGGCAAGCCATTCTGTCCGC 1292  
Db 2021 GTGTTGACGGGACAGGGGGCGGGCAAGGGGCTTCCCGGGGCAAGCCATTCTGTCCGC 2080  
QY 1293 CATGACAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAAGCGCGGGAAGAATC 1352  
Db 2081 CATGACAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAAGCGCGGGAAGAATC 2140  
QY 1353 GATCAAGGACTTACCCTGCGCTGCGGGAAGGTGTGACCCGTCGACCTGGGCA 1412  
Db 2141 GATCAAGGACTTACCCTGCGCTGCGGGAAGGTGTGACCCGTCGACCTGGGCA 2200  
QY 1413 GCACGAAGCAGCTGCAACGTCACATGGCGGCACTGCAAGTTCTGTGGGAAGCAT 1472  
Db 2201 GCACGAAGCAGCTGCAACGTCACATGGCGGCACTGCAAGTTCTGTGGGAAGCAT 2260  
QY 1473 CTTCAACAGATCAGGACGCGGAGGCGAGCAAGAGAGCTTGGCGCTTCTGTGGGAAGC 1532  
Db 2261 CTTCAACAGATCAGGACGCGGAGGCGAGCAAGAGAGCTTGGCGCTTCTGTGGGAAGC 2320  
QY 1533 CGCCGACGCTTCGGCAAGCTCAGGCGGCTGAGAGAGCGGGCAAGCGGCAACCT 1592  
Db 2321 CGCCGACGCTTCGGCAAGCTCAGGCGGCTGAGAGAGCGGGCAAGCGGCAACCT 2380  
QY 1593 TGTTCGAGCGCGCGCGAGCGCCCTGAACGCGCTTGAAGAGCTGTACGAAGACCGCGGCG 1652  
Db 2381 TGTTCGAGCGCGCGCGAGCGCCCTGAACGCGCTTGAAGAGCTGTACGAAGACCGCGGCG 2440  
QY 1653 AGGAGCTTACGAGCGGACCCGTTGGCAGGAAGACACTTCCGGAAGCAACAGGACGCGTGAC 1712  
Db 2441 AGGAGCTTACGAGCGGACCCGTTGGCAGGAAGCAACTTCCGGAAGCAACAGGACGCGTGAC 2500

QY 1713 GCTCCGGCAGCAAGGGGCGGAAGAGCGGCTTGCCGAACCTTGAAGCCCGGAAGCCCCGAA 1772  
Db 2501 GCTCCGGCAGCAAGGGGCGGAAGAGCGGCTTGCCGAACCTTGAAGCCCGGAAGCCCCGAA 2560  
QY 1773 GTTGCCCTTGAACCAATGTTCCCGGAAGACCGGACGCTGACCGGCGCTAAGTC 1832  
Db 2561 GCTTCCCTTGAACCAATGTTCCCGGAAGACCGGACGCTGACCGGCGCTAAGTC 2620  
QY 1833 GTGTGGGGGCGCGCTCAGTAGACGACAAAGCGGCTGTCGTGGGCTCTTGTAGACAA 1892  
Db 2621 GTGTGGGGGCGCGCTCAGTAGACGACAAAGCGGCTGTCGTGGGCTCTTGTAGACAA 2680  
QY 1893 GATCGTTGTACGAAGTGCAGTACGGGACGGGGGACGGAACCCCATCGAGAAGCGGC 1952  
Db 2681 GATCGTTGTACGAAGTGCAGTACGGGACGGGGGACGGAACCCCATCGAGAAGCGGC 2740  
QY 1953 TTCGATCAGTGGGGCGAAGCCCGGACGACGACGAGCAAGACGACGCCAGGACGCGAC 2012  
Db 2741 TTCGATCAGTGGGGCGAAGCCCGGACGACGACGAGCAAGACGACGCCAGGACGCGAC 2800  
QY 2013 GGAAGAGCTAGCGGCGTAG 2031  
Db 2801 GGAAGAGCTAGCGGCGTAG 2819  
RESULT 6  
PCL414670  
LOCUS  
DEFINITION Cloning vector PSET152.  
ACCESSION AJ414670  
VERSION AJ414670.1 GI:17974209  
KEYWORDS aac(3)IV gene; apramycin resistance gene; int gene; integrase; lacZ  
SOURCE alpha peptide; lacza gene.  
ORGANISM Cloning vector PSET152  
REFERENCE 1  
AUTHORS Wilkinson,C.J., Hughes-Thomas,Z.A., Martin (nee Rowe),C.J.,  
Bohm,I., Milronenko,T., Deacon,M., Wheatcroft,M., Wirtz,G.,  
Staunton,J. and Leadlay,P.F.  
TITLE Increasing the efficiency of heterologous promoters in  
actinomycetes  
JOURNAL J. Mol. Microbiol. Biotechnol.  
REFERENCE 2 (bases 1 to 5715)  
AUTHORS Wilkinson,C.J.  
TITLE Direct Submission  
JOURNAL Submitted (28-SEP-2001) Wilkinson C.J., Department of Biochemistry,  
University of Cambridge, 80 Tennis Court Rd., Cambridge, CB2 1GA,  
UNITED KINGDOM  
COMMENT Related sequences V00635, J02451, M24524, M21778.  
FEATURES  
source  
1..5715  
/organism="Cloning vector PSET152"  
/specific\_host="Streptomyces sp."  
/db\_xref="taxon:173260"  
/lab\_host="Escherichia coli"  
complement(163..507)  
/gene="lacza"  
complement(163..507)  
/gene="lacza"  
complement(163..507)  
/function="blue-white screening"  
/codon\_start=1  
/transl\_table=11  
/product="lacZ alpha peptide"  
/protein\_id="CAC93946.1"  
/db\_xref="GI:17974210"  
/translation="MTNSISRAAADPLESTCSPLALAVLORDMENPGVQLNRL  
AAHPFASWRNSEARTRPSQQLRSLNGEWRIMRYFLTLGISHRINSPMSSTSG  
IGSAADAKCRST"  
complement(508..509)  
/note="pUC18"  
/replace="ggt"  
variation



```
gene      1729. .2505
          /gene="aac(3)IV"
CDS       1729. .2505
          /gene="aac(3)IV"
          /function="apramycin resistance"
          /codon_start=1
          /transl_table=11
          /product="apramycin resistance gene"
          /protein_id="CAC93947.1"
          /db_xref="GI:17974212"
          /translation="MQEWRKAEILGQLNLGVTGPGVLLVHSSRSVRPLEDPLGL
          IEALRAALGPGGLVMPWSGLEDDEDPATSPATPDIGVSDTFWRLPNYKRSAPF
          AFAAGPQAEQIISDPLPLPHSPASPVARVHELDOVLLGVHDANTLHLAELMA
          KVPYGVPRHCTILODGKLVRYDYLENDHCERFALADRWLKEKSLQKEGPGVGHAFARL
          IRSRDIVATALGQLGRDPLIFLHPPEAGCEECDAARQSIG"
          3340. .3451
          /note="RP4 origin of single-stranded DNA transfer"
          /direction=RIGHT
          3799. .3837
          /note="phic31 attp site"
          3854. .5695
          /gene="int"
          3854. .5695
          /gene="int"
          /function="phic31 integrase"
          /codon_start=1
          /transl_table=11
          /product="Integrase"
          /protein_id="CAC93948.1"
          /db_xref="GI:17974212"
          /translation="MTQGVVTGVDTYAGAYDRSRENERSSAASPAIORSANEDKAD
          LQREVERDGRFRFVGHFSEAPGTSAGTAERPEFERILNECRAGRNMIIIVDVSF
          SRLKVDALPIVSELLALGTVIVSTOEGVFROGNVMDLHILMLDASHKESLSKSAK
          ILDTKNLQRELGGYVGKAPYGEELVSEKTEIRNGRMNVNINLASHSTPLTGPFE
          EEPDVIRWWRREIKTHKHLPEKPSQAAIHPSITGLCKRMDADAVPRGETIGKTA
          SSAMDPAIVMRILRDPRIAGFAAEVIVYKKKPDGTPPTKIEGRIQRPITLRVELDC
          GPIEPAEWYELQAWLDEGRGKLSRGQALISAMDLYCECGAVMTSKGEESIKDS
          YRCRRKVVDPSPAGQHEGTCNSMALDKFYAEIRFNKIRHAEGBDETLALLWEAR
          RRGKLEAPEKSGERANLVAERADALNALIELEDRAAGAYDGPVGRKHKRQQAALT
          LROGAEEERLAELEAEPKLPIDOWFPEADADPTGPKSWMGASVDDKRVFVGLFV
          DKIVVTKSTTGRGQGTPEKRSITWAKPPDDDEDADAODGTEDVAA"
```

BASE COUNT 1189 a 1643 c 1745 g 1138 t  
ORIGIN

Query Match 88.5%; Score 1796.6; DB 12; Length 5715;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 1805; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```
QY 213 GATGTGAGGTACGCGGGTGTCTTACGACCGTCACTGCGCGAGCGGAGATAGCAGTGC 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3877 GGTGACACGTAAGCGGGTGTCTTACGACCGTCACTGCGCGAGCGGAGATTCGAGCGC 3936

QY 273 AGCAAGCCGACGACACAGCGTAGCGCCCAAGAGACAGAGCGCGCGACCTTCAGCGCGA 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3937 AGCAAGCCGACGACACAGCGTAGCGCCCAAGAGACAGAGCGCGCGACCTTCAGCGCGA 3996

QY 333 AGTCAGCGCGAGCGGGGGCGGTTGCTGCGGCAATTCAGCGAAGCGCGCGGCGAC 392
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3997 AGTCAGCGCGAGCGGGGGCGGTTGCTGCGGCAATTCAGCGAAGCGCGGCGAC 4056

QY 393 GTCGCGCTTCGGGAGCGCGGAGCGCGCGGAGTTCGAGCGCATCTGAACGATGCCGCGC 452
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4057 GTCGCGCTTCGGGAGCGCGGAGCGCGCGGAGTTCGAGCGCATCTGAACGATGCCGCGC 4116

QY 453 CGGCGCGCTCAACATGATGATGCTATGAGCTGTGCGGCTTCTGCGCGCTGAAGGTCA 512
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4117 CGGCGCGCTCAACATGATGATGCTATGAGCTGTGCGGCTTCTGCGCGCTGAAGGTCA 4176

QY 513 GGACGCGATTCGATGCTCGGAATGCTCGCCCTGCGCGGTGACGATGTTTCCACTCA 572
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4177 GGACGCGATTCGATGCTCGGAATGCTCGCCCTGCGCGGTGACGATGTTTCCACTCA 4236

QY 573 GGAAGCGCTCTTCGCGCAGGGAACGTCATGACCTGATTCACCTGATATGCGGCTCGA 632
```

```
Db 4237 GGAAGCGCTCTTCGCGCAGGGAACGTCATGACCTGATTCACCTGATATGCGGCTCGA 4296

QY 633 CGCGTCGACAAAGATCTTCGCTGAAGTCGGCGAAGATTCTCGACACGAGAACCCTTCA 692
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4297 CGCGTCGACAAAGATCTTCGCTGAAGTCGGCGAAGATTCTCGACACGAGAACCCTTCA 4356

QY 693 GCGCGAATGGCGGGTACGTCGCGCGGGAAGGCGCCTTACGCGCTTCGAGCTTGTTCGA 752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4357 GCGCGAATGGCGGGTACGTCGCGCGGGAAGGCGCCTTACGCGCTTCGAGCTTGTTCGA 4416

QY 753 GACGAAGGATTCACGCGCAACGCGCGAATGCTCAATGTCGTATCAACAAGTTAGCGCA 812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4417 GACGAAGGATTCACGCGCAACGCGCGAATGCTCAATGTCGTATCAACAAGTTAGCGCA 4476

QY 813 CTCGACCACCTCCCTTACCGGACCCCTTCGAGTTTCGAGCCCGACGTAATCCGCTGTGCTG 872
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4477 CTCGACCACCTCCCTTACCGGACCCCTTCGAGTTTCGAGCCCGACGTAATCCGCTGTGCTG 4536

QY 873 GCGTGAGATCAAGACGACACAACACCTTCCCTTCAAGCCGCGGACGTCAGCCGCCATTCA 932
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4537 GCGTGAGATCAAGACGACACAACACCTTCCCTTCAAGCCGCGGACGTCAGCCGCCATTCA 4596

QY 933 CCGCGGACGATCAACGGGGCTTTGTAAGCGCATGAGACGCTGACGCCGTGCCAGCCCGGG 992
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4597 CCGCGGACGATCAACGGGGCTTTGTAAGCGCATGAGACGCTGACGCCGTGCCAGCCCGGG 4656

QY 993 CGAGACGATTGGGAAGAAGACCCGCTTCAAGCCGCTGGAGCCCGGCAACCGTTATGCGAAT 1052
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4657 CGAGACGATTGGGAAGAAGACCCGCTTCAAGCCGCTGGAGCCCGGCAACCGTTATGCGAAT 4716

QY 1053 CCTTCGGGACCCCGCTATTGCGGGCTTCGCGGCTGAGGTGATCTACAAGAAGACCCGGA 1112
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4717 CCTTCGGGACCCCGCTATTGCGGGCTTCGCGGCTGAGGTGATCTACAAGAAGACCCGGA 4776

QY 1113 CGGACGCGGACCAAGAGATTGAGGGTTACCGCATTCAGCGGACCCGATACGCTCCG 1172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4777 CGGACGCGGACCAAGAGATTGAGGGTTACCGCATTCAGCGGACCCGATACGCTCCG 4836

QY 1173 GCGGCTGACGCTTGTATGCGGACCGATCATGAGCCCGCTGAGTGTATGAGCTTCAGGC 1232
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4837 GCGGCTGACGCTTGTATGCGGACCGATCATGAGCCCGCTGAGTGTATGAGCTTCAGGC 4896

QY 1233 GTGTTGAGCGGACGAGGGGGCGGCGCAAGGGCTTTCGCGGGGCAAGCATTCTGTCCGC 1292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4897 GTGTTGAGCGGACGAGGGGGCGGCGCAAGGGCTTTCGCGGGGCAAGCATTCTGTCCGC 4956

QY 1293 CATGACAAGCTGTACTGCGAGTGTGGCGCCGTCACTTCGAAGCGCGGGGAAGAATC 1352
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4957 CATGACAAGCTGTACTGCGAGTGTGGCGCCGTCACTTCGAAGCGCGGGGAAGAATC 5016

QY 1353 GATCAAGGACTCTTACCGCTGCCGTCGCCGGAAGGTGTCGACCCGCTCCGCACTGGGCA 1412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5017 GATCAAGGACTCTTACCGCTGCCGTCGCCGGAAGGTGTCGACCCGCTCCGCACTGGGCA 5076

QY 1413 GCACGAAGCAGTGCACAGCTACGATGGCGGCACTGCACAAGTTGCTGCGGAACGCAT 1472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5077 GCACGAAGCAGTGCACAGCTACGATGGCGGCACTGCACAAGTTGCTGCGGAACGCAT 5136

QY 1473 CTTCAACAAGATCAGGACGCGCGAAGGCGACGACGAGAGAGAGCTTGGCGCTTCTGTGGAAGC 1532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5137 CTTCAACAAGATCAGGACGCGCGAAGGCGACGACGAGAGAGAGCTTGGCGCTTCTGTGGAAGC 5196

QY 1533 CGCCGACGCTTCGGCAAGCTCACTGAGGCGCTGAGAAGAGCGGCGCAACGGCGAACC 1592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5197 CGCCGACGCTTCGGCAAGCTCACTGAGGCGCTGAGAAGAGCGGCGCAACGGCGAACC 5256

QY 1593 TGTTCGGAGCGCGCGCAAGCCCTGAACGCGCTTGAAGAGCTGTACGAAGACCGCGCGC 1652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5257 TGTTCGGAGCGCGCGCAAGCCCTGAACGCGCTTGAAGAGCTGTACGAAGACCGCGCGC 5316

QY 1653 AGGAGCTTACGACGACCCGCTTGCGAGGAAGCACTTCGGAAGCAACGACGCGCTGAC 1712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 5317 AGGCGGTACGACGAGCCCGTTGGCAGGAAGCACATTCCGGGAAGCAACAGCGACCGCTGAC 5376  
Oy 1713 GCTCCGGCAGCAAGGGGGCGAAGACGGCTTGCCGAATTGAAGCCGCGGAGCCCCGAA 1772  
Db 5377 GCTCCGGCAGCAAGGGGGCGAAGACGGCTTGCCGAATTGAAGCCGCGGAGCCCCGAA 5436  
Oy 1773 GTTGGCCCTTGACCAATGTTTCCCCGAAGACGGCCGACGCTGACCCGCGCCCTAAGTC 1832  
Db 5437 GCTTCCCCCTTGACCAATGTTTCCCCGAAGACGGCCGACGCTGACCCGCGCCCTAAGTC 5496  
Oy 1833 GTGGTGGGGGGGGCGCGCTCAGTAGACGACAAAGCCGCTGCTGGGCTCTTGTAGACAA 1892  
Db 5497 GTGGTGGGGGGGGCGCGCTCAGTAGACGACAAAGCCGCTGCTGGGCTCTTGTAGACAA 5556  
Oy 1893 GATCGTTGTACGAAGTCACTACGGGCGAGGGGCGAAGCGCCCATCGAGAAGCGCGC 1952  
Db 5557 GATCGTTGTACGAAGTCACTACGGGCGAGGGGCGAAGCGCCCATCGAGAAGCGCGC 5616  
Oy 1953 TTGCATCAGCTGGGCGAAGCCCGCAGCGACGACGAGCAAGACGACGCCCGCAGCGGCAC 2012  
Db 5617 TTGCATCAGCTGGGCGAAGCCCGCAGCGACGACGAGCAAGACGACGCCCGCAGCGGCAC 5676  
\*Oy 2013 GGAAGACGTAGCGCGCTAG 2031  
Db 5677 GGAAGACGTAGCGCGCTAG 5695

RESULT 7  
BPH6589 41489 bp DNA linear PHG 12-MAY-2000  
LOCUS BPH6589  
DEFINITION Bacteriophage phi-C31 complete genome.  
ACCESSION AJ006589  
VERSION AJ006589.1 GI:3947449  
KEYWORDS C gene; dCMP deaminase; deoxynucleotide monophosphate kinase; DNA polymerase; int gene; integrase; large subunit; major capsid protein; major tail protein; portal protein; protease; putative; repressor; tail fibre protein; tail tape measure protein; terminase; transfer RNA-Thr; tRNA-Thr.

SOURCE Bacteriophage phi-C31.  
ORGANISM Bacteriophage phi-C31.  
viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae; Lambda-like viruses.

REFERENCE 1 (bases 1 to 41489)  
AUTHORS Hendrix,R.W., Smith,M.C., Burns,R.N., Ford,M.E. and Hatfull,G.F.  
TITLE Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (5), 2192-2197 (1999)

JOURNAL MEDLINE 99162580  
PUBMED 10051617  
REFERENCE 2 (bases 1 to 41489)  
AUTHORS Smith,M.C., Burns,R.N., Wilson,S.E. and Gregory,M.A.  
TITLE The complete genome sequence of the Streptomyces temperate phage straight phic31: evolutionary relationships to other viruses  
JOURNAL Nucleic Acids Res. 27 (10), 2145-2155 (1999)

JOURNAL MEDLINE 99238410  
PUBMED 10219087  
REFERENCE 3 (bases 1 to 41489)  
AUTHORS Smith,M.C.M.  
TITLE Direct Submission  
JOURNAL Submitted (01-JUN-1998) Smith M.C.M., Genetics, University of Nottingham, Queens Medical Centre, Nottingham, NG7 2UH, UK  
FEATURES  
source  
1. 41489  
/organism="Bacteriophage phi-C31"  
/vifion  
/strain="Norwich stock"  
/db\_xref="taxon:10719"  
145. 585  
/gene="31"  
145. 585  
/gene="31"  
/codon\_start=1  
/transl\_table=11

gene  
CDS

/product="gp31"  
/protein\_id="CAA07101.1"  
/db\_xref="GI:3947450"  
/db\_xref="SPTREMBL:Q37836"  
/translation="MSRAKSPELRTGNANAAESAAPVVEGRAPVPAHLKATGKDY  
WRNVWAGMGAYSPTDRNVILRYCELHDBRADLSLEADGYMSEGYNGOPVAHPML  
RYVESTKEKLSIERIAIGFTPEARMRLGIYAAEARKVAAGPEDF"  
complement(640..1170)  
/gene="32"  
complement(640..1170)  
/gene="32"  
/note="transcribed in lysogens"  
/codon\_start=1  
/transl\_table=11  
/product="gp32"  
/protein\_id="CAA07102.1"  
/db\_xref="GI:3947451"  
/db\_xref="SPTREMBL:Q37837"  
/translation="MNAATPTRYFGKTFNGLHYVRVLRGEELYRTPGYTTDMAVAD  
AACWEAFRVVAPAEETPAEKDRMKRDEAGVDKPEVGLREGDYVHTTYTMKAATNRKG  
ORVSEYOVSEWECSGAMARDWARGLYRVLVGFTEGAYANKGEIVKVRDPKALKR  
AEAEATRAVYAESKR"  
1282..2841  
/gene="33"  
1282..2841  
/gene="33"  
/function="putative large terminase subunit"  
/codon\_start=1  
/transl\_table=11  
/product="gp33"  
/protein\_id="CAA07103.1"  
/db\_xref="GI:3947452"  
/translation="MADWAGIDPVIARHIPADATVPSEGYRVAKWIEFCYLTGSPAG  
QPFRLPWORELIDAVVLTQDTFGWRMRKRRTVVVCVARKNGKSTIAAIMLYHLIA  
DRGDAQROIIAAANDRNQARVFDSSAKQVNASPKLAACDVQRIYRKDNTYRVVS  
ADAGROOGLNPAASLDEYAFSKHSDLPDALTLGSAARNQPMFLISTAGPDPGFA  
ALCEOGERVNSGEADDPTELYRSGMPKLGFTVDLDPDVWRACNSYDILNPDDFKAA  
AORSTEASFRIYRLSQFVRGASTWLPFGLDLSLAADDPLEPGDEVVLGFDGSKGDS  
TALVACRIRDLKVFVLGWEAPADADAHWRVPMADVREELHTALDYRVRLVADPYRM  
EETLDNLEADGFPVEAFPTNSLARMPATQAVYDACRDGRLSHDCNPALGRHIGNAVL  
KEDARGARITKEHASSRRKIDLAVAMVAVHGAVMWREDNGIVSDKPIIATWEDDEGN  
VFVHPDHAEEF"  
2886..4259  
/gene="34"  
2886..4259  
/gene="34"  
/function="putative portal protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp34"  
/protein\_id="CAA07104.1"  
/db\_xref="GI:3947453"  
/db\_xref="SPTREMBL:Q9ZXB2"  
/translation="MGFWSALFGRGHSPALDGEIARAWEPYDPSIYNLGAVAASGETV  
TPHDALQVSASFASVRLSETIATPLSTYSKRGSRKEIVTPENDIYPNAEPGGMGR  
IDILSOTVLSLLQGNAPLAVRWQGPNIYGLDYLPTKIHVHMVAVDGLRKKVFEAYD  
IDADGEVLLGWFTPRDVLHIPGMPLGDFVGCSPISYARESIGLAAOKYGSKFA  
NGAMPGAVEVPGTSEEGFLARAREAWRAANSQVDNAHRVALLTEGAKFSKVAMSPDE  
AQFLQTRQFQVPEIARIFEGVPHLISDATNSTWSGSLAEONIAFMESLRPWLRIE  
AGFNRLFAETADRRFVKFNLDIEIKRGAPKERMELWSLGLONGIYSIDEVRAEDMT  
PLPDGLGEKYRVPLNLGVEGEEPEPEPAPPAIEPPAEEDFEPEEGKPDDEGATE  
EDDEDDA"  
4240..4881  
/gene="35"  
4240..4881  
/gene="35"  
/note="putative protease"  
/codon\_start=1  
/transl\_table=11  
/product="gp35"  
/protein\_id="CAA07105.1"  
/db\_xref="GI:3947454"  
/db\_xref="SPTREMBL:Q9ZXB1"

gene  
CDS

gene  
CDS

gene /translation="MTKTTRELRFVAVGTLEERASDDGRISMRYAYRENELSHDLGFE  
RERIVPGAGAPSLRONDVYATFNHDTSLIGRTSGTLRVGEDREGWEIDLPDPTV  
GRDVAKLKRGDLOGSSFTFRVLDGGQRADDDPETGLPVREITAMDVVELGPPVNP  
AYPTQASLSRISEETLRIGEFAPPTTEERDSQPDGDVAPASIHFDARSLVALSK"  
4893. .6071  
/gene="36"  
4893. .6071  
/gene="36"  
/function="major capsid protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp36"  
/protein\_id="CAA07106.1"  
/db\_xref="GI:3947455"  
/db\_xref="SPTREMBL:Q9ZMV6"  
/translation="MDATILSANFEARERATAELRSLTDEFACKEMTAAREKEERLL  
TAVADFDRIRKIDAIKATDAVTSLSGSGSGAQRSAHDHDDAVLRAGNLGEAR  
SFERAPEKRDGTAKAGNPVLSRTLYQLIAQAYERSAIMRGASTFTSDANPMDFTV  
ITGRATAGIVGETAEIPESYPATQSRMGFKYGFASVSYEFATDQVLDLVGLVSD  
AGPAIGDAMGRHFLTGTGQPRGILTDAVGANAAFGADADSKVSDALIDLFHEVPS  
AVRKNKAEVNDLRAQMRKLDKANGQYLMQSALTGPADTFNGKVVETDDGMPADKV  
LEADLSKYRVRFAGSLRVDKSTDIYRFLQADGLLVARGAKVLTVTTPAA"  
mat\_peptide 5226. .6068  
/gene="36"  
/product="gp36"  
6144. .6512  
/gene="37"  
6144. .6512  
/gene="37"  
/codon\_start=1  
/transl\_table=11  
/product="gp37"  
/protein\_id="CAA07107.1"  
/db\_xref="GI:3947456"  
/db\_xref="SPTREMBL:Q9ZXB0"  
/translation="MAVATIELRALDGLDSDALFSDLLSDAIDFSVETVEAYCGRK  
MDTAEPTPETIRMCVRLARQVLDHVSRIIPDRALQLQSEFGSIQLAQAGGNWRPPT  
LPEVNAKLNLVRLPLFIEM"  
6519. .6956  
/gene="38"  
6519. .6956  
/gene="38"  
/codon\_start=1  
/transl\_table=11  
/product="gp38"  
/protein\_id="CAA07108.1"  
/db\_xref="GI:3947457"  
/db\_xref="SPTREMBL:Q9ZXA9"  
/translation="MALIFNAKVRLEALKANVPGDVQCTFAETGDNRRKQVWLIGAT  
VDDDLAPVAMRSGAKPTNVTGYVEAHAVTTPGNPVDARAVYGIIDYVKAACAANA  
DLVSVPLMDVREPESASVESTETTDGAYSALTVRVVRGRVYQ"  
6989. .7975  
/gene="39"  
6989. .7975  
/gene="39"  
/function="major tail protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp39"  
/protein\_id="CAA07109.1"  
/db\_xref="GI:3947458"  
/db\_xref="SPTREMBL:Q9ZXAB"  
/translation="MALDASIGIGREDTYGTLGSGVEGEGQADSMKQTVEPISVGF  
RAGMQTARADRRNMVNMGELELDVLDAGAGSLTFAFDKVTVDTGKVTYLET  
SDVSEAPSTAQMVPRGVDGSKVAFKHLGCVATEWSLTAVEEAVKLAFTDFRDATH  
TSTPAEIVEPTYPAEAYPYDWTRTAVELSRAGSVVEFDATSLTELGELGMKIDRRFLN  
GSELKKPYRNALPTYEGLTEGERSAASLGLYDAFVSGEYCSFKATFFGILPGSSLSV  
EAPAIQFTGESPEAATDEVTIHNPFRVLDPGDGVTPAKILTYVEPTPEP"  
7975. .8406  
/gene="40"  
7975. .8406  
/gene="40"  
CDS /gene="40"

Query Match									
Best Local Similarity 99.2%; Pred. No. 0;									
Matches 1805; Conservative 0; Mismatches 14; Indels 0; Gaps 0;									
gene									
/codon_start=1									
/transl_table=11									
/product="gp40"									
/protein_id="CAA07110.1"									
/db_xref="GI:3947459"									
/db_xref="SPTREMBL:Q9ZXA7"									
/translation="MAQRSAYTIQVDGLRQFQNRVRLRDKELNKAVREANKASGEVL YPOAKHESPDGHRDPRSSKRYRPGKLDKSIKVTASAGAVIKAGSARVPYAAAIHFG YRKNISANRFLYRAMARKSDVVAATYERTIAAVEKYLEY"									
8406. .8878									
QY	213	GATGTGAGGTACGGCGGTGCTTACGACCGTCACTCGCGGAGCGCGAGAATAGCAGTGC	272						
DB	38444	GGTGACACGTACGGCGGTGCTTACGACCGTCACTCGCGGAGCGCGAGAATTCAGCGCC	38503						
QY	273	AGCAAGCCGACGACACAGCGTACGCCCAACGAGACAGCGCGCGACCTTCACCGCGA	332						
DB	38504	AGCAAGCCGACGACACAGCGTACGCCCAACGAGACAGCGCGCGACCTTCACCGCGA	38563						
QY	333	AGTCGAGCGGACGCGCGCGGCTTACAGTTCGTCGGGCATTTACGGAAGCGCGCGCAC	392						
DB	38564	AGTCGAGCGGACGCGCGCGGCTTACAGTTCGTCGGGCATTTACGGAAGCGCGCGCAC	38623						
QY	393	GTCGCGCTTCGGGACGCGCGGCGCGCGAGTTCGACAGCATCTGAAACGAAATGCCGCGC	452						
DB	38624	GTCGCGCTTCGGGACGCGCGGCGCGCGAGTTCGACAGCATCTGAAACGAAATGCCGCGC	38683						
QY	453	CGGCGGCTCAACATGATCATGTCTATGACGTGTCGCGCTTCTCGCGCTGAAGGTCAT	512						
DB	38684	CGGCGGCTCAACATGATCATGTCTATGACGTGTCGCGCTTCTCGCGCTGAAGGTCAT	38743						
QY	513	GGACGCGATTCGATTTGTCGGAATTGCTCGCGCTGGCGGTGACGATGTTTCCACTCA	572						
DB	38744	GGACGCGATTCGATTTGTCGGAATTGCTCGCGCTGGCGGTGACGATGTTTCCACTCA	38803						
QY	573	GGAAGCGCTCTCCGGCAGGGAACGTATGACCTGATTCACCTGATTAATGCGGCTCGA	632						
DB	38804	GGAAGCGCTCTCCGGCAGGGAACGTATGACCTGATTCACCTGATTAATGCGGCTCGA	38863						
QY	633	CGCGTCGACAAAGAAATCTTCGCTGAAGTCGCGGAGATTTCTGACACGAAACCTTCA	692						
DB	38864	CGCGTCGACAAAGAAATCTTCGCTGAAGTCGCGGAGATTTCTGACACGAAACCTTCA	38923						
QY	693	GCGCGAATTGGCGGCTACGTCGGCGGGAAGCGCGCTTACGGCTTCGAGCTTGTTCGA	752						
DB	38924	GCGCGAATTGGCGGCTACGTCGGCGGGAAGCGCGCTTACGGCTTGTTCGA	38983						
QY	753	GACGAAGGAGATCACGCGGACGCGCGAATGTTCAATGTCGTCATCAACAAGTTAGCGCA	812						
DB	38984	GACGAAGGAGATCACGCGGACGCGCGAATGTTCAATGTCGTCATCAACAAGTTAGCGCA	39043						
QY	813	CTCGACCACTCCCTTACCGGACCTTCGAGTTCGAGCCGCGACGTAATCCGGTGTGTG	872						
DB	39044	CTCGACCACTCCCTTACCGGACCTTCGAGTTCGAGCCGCGACGTAATCCGGTGTGTG	39103						
QY	873	GCGTGAGATCAAGACGCAACAACACCTTCCCTTCAAGCCGGGCGAGTCAAGCCGCATTC	932						
DB	39104	GCGTGAGATCAAGACGCAACAACACCTTCCCTTCAAGCCGGGCGAGTCAAGCCGCATTC	39163						
QY	933	CCCGGGCAGCATCACGGGCGCTTGTAAAGCGCATGAGCGCTGACGCGCTGCGACCCGGGG	992						
DB	39164	CCCGGGCAGCATCACGGGCGCTTGTAAAGCGCATGAGCGCTGACGCGCTGCGACCCGGGG	39223						
QY	993	CGAGACGATTGGGAAGAAGACCGCTTCAAGCGCTTGGAGACCCGCGCAACCGTTATGCGAAT	1052						
DB	39224	CGAGACGATTGGGAAGAAGACCGCTTCAAGCGCTTGGAGACCCGCGCAACCGTTATGCGAAT	39283						
QY	1053	CCTTCGGGACCCGCGCTATTCGGGCTTCGCGCTGAGGTGATCTACAGAAGAACCCGGA	1112						

Db 39284 CCTTCGGACCCCGGTATTGCGGCTTCGCCGTGAGTGATCTACAAGAAGACCCGGA 39343  
Oy 1113 CGGCACCCGACACGAGATTGAGGGTTACCGCATTCAGCGCGACCGATCAGCGTCCG 1172  
Db 39344 CGGCACCCGACACGAGATTGAGGGTTACCGCATTCAGCGCGACCGATCAGCGTCCG 39403  
Oy 1173 GCCGGTCGAGCTTGATTGCGGACCGATCATCGACCCCGCTGAGTGTATGAGCTTCAGGC 1232  
Db 39404 GCCGGTCGAGCTTGATTGCGGACCGATCATCGACCCCGCTGAGTGTATGAGCTTCAGGC 39463  
Oy 1233 GTGTTGACGCGCAGGGGGCGCGCAAGGGGCTTCCCGGGGCAAGCCATTCTGTCCGC 1292  
Db 39464 GTGTTGACGCGCAGGGGGCGCGCAAGGGGCTTCCCGGGGCAAGCCATTCTGTCCGC 39523  
Oy 1293 CATGACAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAAGCGCGGGAAGATC 1352  
Db 39524 CATGACAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAAGCGCGGGAAGATC 39583  
Oy 1353 GATCAAGACTCTTACCGCTGCGCGCGGAAGGTGTCGACCCGTCGCACCTGGGCA 1412  
Db 39584 GATCAAGACTCTTACCGCTGCGCGCGGAAGGTGTCGACCCGTCGCACCTGGGCA 39643  
Oy 1413 GCACGAAGCAGCTGCAACGTACAGATGGCGGCACTGCAAGTTCGTTGCGGAAGCAT 1472  
Db 39644 GCACGAAGCAGCTGCAACGTACAGATGGCGGCACTGCAAGTTCGTTGCGGAAGCAT 39703  
Oy 1473 CTTCAACAAGATCAGGCACGCGCGAAGCGCAGAGAGACTTGGCGCTTCTGTGGAAGC 1532  
Db 39704 CTTCAACAAGATCAGGCACGCGCGAAGCGCAGAGAGACTTGGCGCTTCTGTGGAAGC 39763  
Oy 1533 CGCCCGACGCTTCGGCAAGCTCCTGAGAGCGCGCTGAGAAGCGCGGCAACGGCGCAACT 1592  
Db 39764 CGCCCGACGCTTCGGCAAGCTCCTGAGAGCGCGCTGAGAAGCGCGGCAACGGCGCAACT 39823  
Oy 1593 TGTTCGCGAGCGCGCGCGAGCGCCCTGAGAGCGCTGTACGAAGCGCGCGGC 1652  
Db 39824 TGTTCGCGAGCGCGCGCGAGCGCCCTGAGAGCGCTGTACGAAGCGCGCGGC 39883  
Oy 1653 AGGAGCTTACGACGAGCGCCGTTGGCGAAGCACTTCGGAAGCAACAGGACGCGTGCAC 1712  
Db 39884 AGGAGCTTACGACGAGCGCCGTTGGCGAAGCACTTCGGAAGCAACAGGACGCGTGCAC 39943  
Oy 1713 GCTCCGCGAAGAGGGGCGGAAGAGGGGCTTGCAGACTTGAAGCGCGCGAAGCCCGCAA 1772  
Db 39944 GCTCCGCGAAGAGGGGCGGAAGAGGGGCTTGCAGACTTGAAGCGCGCGAAGCCCGCAA 40003  
Oy 1773 GTTCCCCCTTGACCAATGTTCCCGAAGACGCGCGTGAACCGCGCCCTTAAGTC 1832  
Db 40004 GCTCCCCCTTGACCAATGTTCCCGAAGACGCGCGTGAACCGCGCCCTTAAGTC 40063  
Oy 1833 GTGTGGGGGCGCGCTCAGTAGACGACAGCGCGTGTGCGGGCTCTTCTAGACAA 1892  
Db 40064 GTGTGGGGGCGCGCTCAGTAGACGACAGCGCGTGTGCGGGCTCTTCTAGACAA 40123  
Oy 1893 GATCGTTGACGAAGTGCAGTACGCGGAGGGGAGGAGCGCCCATCGAAGCGCGC 1952  
Db 40124 GATCGTTGACGAAGTGCAGTACGCGGAGGGGAGGAGCGCCCATCGAAGCGCGC 40183  
Oy 1953 TTGATCAGCTGGGCGAAGCGCGGACCGACGACGACGAGAGAGAGCGCCAGAGCGGAC 2012  
Db 40184 TTGATCAGCTGGGCGAAGCGCGGACCGACGACGACGAGAGAGAGCGCCAGAGCGGAC 40243  
Oy 2013 GGAAGACGTAGCGGCTAG 2031  
Db 40244 GGAAGACGTAGCGGCTAG 40262

RESULT 8  
LOCUS EVU84006 5967 bp DNA circular SYN 03-MAY-1997  
DEFINITION Expression vector pBSII-LUCINT firefly luciferase (LUCINT),  
beta-galactosidase (lacZ) and beta-lactamase (ampR) genes, complete  
cds and lac operon promoter sequence.

ACCESSION U84006  
VERSION U84006.1 GI:2071944  
KEYWORDS  
SOURCE Expression vector pBSII-LUCINT.  
ORGANISM Expression vector pBSII-LUCINT.  
REFERENCE 1 (bases 1 to 5967)  
AUTHORS Mankin,S.L., Allen,G.C. and Thompson,W.F.  
TITLE Introduction of a plant intron into the Luciferase Gene of Photinus  
pyralis  
JOURNAL Plant Mol. Biol. Rep. (1997) In press  
REFERENCE 2 (bases 1 to 5967)  
AUTHORS Mankin,S.L., Allen,G.C. and Thompson,W.F.  
TITLE Direct Submission  
JOURNAL Submitted (06-JAN-1997) Botany, NC State Univ., Box 7612, Raleigh,  
NC 27695-7612, USA  
FEATURES  
source  
1. 5967  
/organism="Expression vector pBSII-LUCINT"  
/db\_xref="taxon:56405"  
/plasmid="pluk07"  
/note="chimera plasmid consisting of a pBS-type plasmid  
(pluk07, dsDNA) derived from the LUC\* gene (a synthetic  
cDNA gene) by the addition of a plant intron sequence"  
52. 851  
/note="Camv 35S promoter"  
875. 3052  
/gene="LUCINT"  
join(875..1029,1219..2716)  
/gene="LUCINT"  
/EC.number="1.13.12.7"  
/note="luciferin 4-monooxygenase; LUC\* with PIV2 intron"  
/codon\_start=1  
/product="firefly luciferase"  
/protein\_id="AAB53627.1"  
/db\_xref="GI:2071945"  
/translation="MEDAKNIKGPAPFYPLEDGTAGEQLHKAMKRYALVPGTIAFTD  
AHIEVNITYAEYFEMSVRLAEAMKRYGLNTNHRIVCSENSLOFEMPVIGALFIGVAV  
APANDIYNERELNSMNIOPVTVFVSKKGLKILNVOKLPPIOKIIMDSKTDYOG  
EQSMYTFVTSHPGPFNEYDFVPESEFDRDKTIALIMNSSGSTGLPKGVALLPHRTACVR  
FSHARDPIFNGQIIPDPAISVVPFHGFMFTLGYLICGFRVVLVMPREEELFLRS  
LDYKIQSALLVPTLSEFAKSTLIDKYDLSNLHEIASGAPLSKEVGEAVAKRFLP  
GIRGGLTETTSAILITPEGDDKPGAVGKVPFEAKVVDLDGTGLGNQRGELCV  
RGPIMISGYVNNPEATNALIDKDGWLHSGDIAYWDEHFIYDRLKSLIKYGYOVA  
PAELBSILLOHPNIFDAGVAGLPDDDAGELPAAYVLEHGKTMTEKEIYDVASQVTT  
AKRLRGVVFDEVPGKLTGLDKDARKIREILIKAKGKSKL"  
1030. 1218  
/gene="LUCINT"  
/note="PIV2 chimeric intron"  
2786. 3052  
/gene="LUCINT"  
/standard\_name="nopoline synthase 3' region"  
/note="nos"  
2816. 2821  
/gene="LUCINT"  
3146. 3355  
/gene="lacZ"  
3146. 3355  
/gene="lacZ"  
/codon\_start=1  
/transl\_table=11  
/product="beta-galactosidase"  
/protein\_id="AAB53629.1"  
/db\_xref="GI:2071947"  
/translation="MAVVLORDWENPGVTQLNRLAAHPFASWRNSEARTDRPSQO  
LRSLNGEWREIVNVNILLKALNFC"  
3895. 4755  
/gene="ampR"  
3895. 4755  
/gene="ampR"  
/function="ampicillin resistance marker"  
/codon\_start=1  
/transl\_table=11

```
/product="beta-lactamase"
/proteln_id="AAB53628.1"
/db_xref="GI:2071946"
/translation="MSIOHFRVALIPEFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
IELDLNSGKILLESFRRPEERFPAMSTKVLCCGAVLSRIDAGQEOIGRIHYSDNLVE
YSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLTTIGGPKELTAFLHNMGDHVTRL
DRWEPELNEAIPNDERDTMPVAMATTLRKLLTGELTLASRQOLIDWMEADKVAGPL
LRSALPAGWFIADKSGAGERGSRGIIAALGPDKPSRTIVITYTGSQATMDERNRQIA
EIGASLIKHW"
5790.5912
/promoter
/Note="lac operon promoter"
BASE COUNT 1686 a 1349 c 1416 g 1516 t
ORIGIN

Query Match 9.8%; Score 200.4; DB 12; Length 5967;
Best Local Similarity 97.1%; Pred. No. 1.3e-27;
Matches 204; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 30 TACGTAAGTTCTGCTTCTACCTTTGATATATATATATATATATATATATAGTA 89
Db 1027 TACGTAAGTTCTGCTTCTACCTTTGATATATATATATATATATATATAGTA 1086
QY 90 ATATAATATTTCAAAATATTTTTCAAAATAAAGAATGTAGTATATAGCAATTGCTTTT 149
Db 1087 ATATAATATTTCAAAATATTTTTCAAAATAAAGAATGTAGTATATAGCAATTGCTTTT 1146
QY 150 CTGTAGTTTATAGTGTATATTTTAAATTTTAACTTTTCTAATATATGACCAAAATTT 209
Db 1147 CTGTAGTTTATAGTGTATATTTTAAATTTTAACTTTTCTAATATATGACCAAAATTT 1206
QY 210 GTTGATGTGCAGGTACGGGGTGCTTACGA 239
Db 1207 GTTGATGTGCAGGTACGGGAATACTTCGA 1236

RESULT 9
AX259239 912 bp DNA linear PAT 26-OCT-2001
LOCUS
DEFINITION Sequence 37 from Patent WO0173087.
ACCESSION AX259239
VERSION AX259239.1 GI:16508485
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1
AUTHORS Hohn,T., Stavolone,L., de Haan,P.T., Ligon,H.T. and Kononova,M.
TITLE Cestrum yellow leaf curling virus promoters
JOURNAL Patent: WO 0173087-A 37 04-OCT-2001;
Syngenta Participations AG (CH)
FEATURES
source
1.912
/organism="synthetic construct"
/db_xref="taxon:32630"
/Note="artificial sequence synthetic GFP gene with intron"
BASE COUNT 245 a 245 c 240 g 182 t
ORIGIN

Query Match 9.8%; Score 198.6; DB 6; Length 912;
Best Local Similarity 95.8%; Pred. No. 2.1e-27;
Matches 204; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 21 CGGGGTGATACGTAAGTTCTGCTTCTACCTTTGATATATATATATATATATATCATTA 80
Db 265 CCGGAGGCGTACGTAAGTTCTGCTTCTACCTTTGATATATATATATATATATATCATTA 324
QY 81 TTAGTAGTAATATATATTTTTCAAAATAAAGAATGTAGTATATAGCA 140
Db 325 TTAGTAGTAATATATATTTTTCAAAATAAAGAATGTAGTATATAGCA 384
QY 141 ATTGCTTTCTGTAGTTTATAGTGTATATTTTAAATTTTAACTTTTCTAATATATGA 200
Db 385 ATTGCTTTCTGTAGTTTATAGTGTATATTTTAAATTTTAACTTTTCTAATATATGA 444
```

```
QY 201 CCAAATTTGTGATGTGACGAGTACGGCGGTGC 233
Db 445 CCAAATTTGTGATGTGACGAGTACGAGGCGC 477

RESULT 10
AX259244 1577 bp DNA linear PAT 26-OCT-2001
LOCUS
DEFINITION Sequence 42 from Patent WO0173087.
ACCESSION AX259244
VERSION AX259244.1 GI:16508490
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1
AUTHORS Hohn,T., Stavolone,L., de Haan,P.T., Ligon,H.T. and Kononova,M.
TITLE Cestrum yellow leaf curling virus promoters
JOURNAL Patent: WO 0173087-A 42 04-OCT-2001;
Syngenta Participations AG (CH)
FEATURES
source
1.1577
/organism="synthetic construct"
/db_xref="taxon:32630"
/Note="artificial sequence Cmpc-syngFPi-nos expression cassette"
BASE COUNT 450 a 372 c 390 g 365 t

Query Match 9.8%; Score 198.6; DB 6; Length 1577;
Best Local Similarity 95.8%; Pred. No. 2.3e-27;
Matches 204; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 21 CGGGGTGATACGTAAGTTCTGCTTCTACCTTTGATATATATATATATATATATCATTA 80
Db 645 CCGGAGGCGTACGTAAGTTCTGCTTCTACCTTTGATATATATATATATATATATCATTA 704
QY 81 TTAGTAGTAATATATATTTTTCAAAATAAAGAATGTAGTATATAGCA 140
Db 705 TTAGTAGTAATATATATTTTTCAAAATAAAGAATGTAGTATATAGCA 764
QY 141 ATTGCTTTCTGTAGTTTATAGTGTATATTTTAAATTTTAACTTTTCTAATATATGA 200
Db 765 ATTGCTTTCTGTAGTTTATAGTGTATATTTTAAATTTTAACTTTTCTAATATATGA 824
QY 201 CCAAATTTGTGATGTGACGAGTACGGCGGTGC 233
Db 825 CCAAATTTGTGATGTGACGAGTACGAGCGC 857

RESULT 11
AX259242 1618 bp DNA linear PAT 26-OCT-2001
LOCUS
DEFINITION Sequence 40 from Patent WO0173087.
ACCESSION AX259242
VERSION AX259242.1 GI:16508488
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1
AUTHORS Hohn,T., Stavolone,L., de Haan,P.T., Ligon,H.T. and Kononova,M.
TITLE Cestrum yellow leaf curling virus promoters
JOURNAL Patent: WO 0173087-A 40 04-OCT-2001;
Syngenta Participations AG (CH)
FEATURES
source
1.1618
/organism="synthetic construct"
/db_xref="taxon:32630"
/Note="artificial sequence CmpS-syngFPi-nos expression cassette"
BASE COUNT 468 a 383 c 400 g 367 t
```







GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 14:41:33 ; Search time 511.5 Seconds  
(without alignments)  
8941.952 Million cell updates/sec

Title: US-09-940-550A-10

Perfect score: 2031  
Sequence: 1 atggcacaagggtgtgtgac.....cggaaagacgtagcgcgctag 2031

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2031	100.0	2031	24	ABK12556	DNA encoding novel
2	2015	99.2	2031	24	ABK12555	DNA encoding novel
3	2015	99.2	9880	24	ABK12557	Plasmid pBPS EW051
4	1796.6	88.5	5711	22	AAD04935	C31-Int expression
5	1795	88.4	1842	22	AAH74879	Nucleotide sequenc
6	1793.4	88.3	3401	11	AAO06904	Sequence encoding
7	1790.2	88.1	1842	22	AAH74880	Nucleotide sequenc
8	1785.6	87.9	1839	22	AAH74881	Nucleotide sequenc
9	1767	87.0	1833	22	AAH74882	Nucleotide sequenc

10	198.6	9.8	912	22	AAD19829	Green fluorescent
11	198.6	9.8	1577	22	AAD19834	Promoter-reporter
12	198.6	9.8	1618	22	AAD19832	Promoter-reporter
13	198.6	9.8	2943	22	AAD19840	Ubq3(At)-synGFP-N
14	195	9.6	2001	22	AAD19830	Beta-glucuronidase
15	195	9.6	2001	24	ABK15667	Glutathione-S-tran
16	195	9.6	2725	22	AAD19835	Promoter-reporter
17	195	9.6	2730	22	AAD19833	Promoter-reporter
18	195	9.6	4072	22	AAD19841	Ubq3(At)-GIG-NOS t
19	195	9.6	4341	22	AAD19839	zmbu1-GIG-NOS term
20	195	9.6	5534	17	AAT43137	PUM1GIT sequence 1
21	195	9.6	12817	24	ABK15668	Expression vector
22	195	9.6	13274	24	ABK15666	Binary vector PNOV
23	195	9.6	1701	21	AAZ29120	Plasmid DV127 used
24	194	9.6	3336	21	AAZ29121	Plasmid DV132 used
25	194	9.6	3877	21	AAZ29123	Plasmid comprising Ca
26	192	9.5	1294	22	AAD19907	PGI vector based o
27	190.8	9.4	1691	22	AAD19910	Glucosamine-6-phos
28	190	9.4	1034	19	AAV54572	Insecticidal prote
29	181.4	8.9	363	22	AAF32096	Insecticidal prote
30	180.6	8.9	363	22	AAF32095	Insecticidal prote
31	180.6	8.9	363	22	AAF32093	Insecticidal prote
32	180.6	8.9	369	22	AAF32094	Insecticidal prote
33	137.8	6.8	1512	22	AAD19909	PTBN1 vector based
34	137.8	6.8	2708	22	AAD19908	PTBN6 vector based
35	137.4	6.8	1851	22	AAD19912	PGI vector based
36	137.4	6.8	2315	22	AAD19911	Human osteoblast d
37	58.2	2.9	154902	24	ABQ88198	Chemically treated
38	53.6	2.6	7752	24	ABL70223	Human osteoblast d
39	53.6	2.6	159400	24	ABQ88126	Chemically treated
40	53	2.6	6641	24	ABL54335	Human osteoblast d
41	53	2.6	6641	24	ABL32314	Human osteoblast d
42	52.2	2.6	6118	24	ABN80110	Human osteoblast d
43	51.8	2.6	1836	24	ABL55639	Human osteoblast d
44	51.8	2.6	4654	22	AA546780	Human osteoblast d
45	51.8	2.6	4654	24	ABL34222	Human osteoblast d

ALIGNMENTS

RESULT 1	ABK12556	standard; DNA; 2031 BP.
ID	ABK12556;	
AC	ABK12556;	
XX		
DT	18-JUN-2002	(first entry)
XX		
DE	DNA encoding novel phi C31 Integrase, phiC31Int*INT.	
XX		
KW	Phi C31 Integrase; recombinase; transgenic; plant;	
KW	agricultural food product; self-excising polynucleotide;	
KW	gene; ds; phiC31Int*INT.	
XX		
OS	Synthetic.	
XX		
PN	WO200216609-A2.	
XX		
PD	28-FEB-2002.	
XX		
PF	27-AUG-2001; 2001WO-US26738.	
XX		
PR	25-AUG-2000; 2000US-227961P.	
XX		
PA	(BADI ) BASF PLANT SCI GMBH.	
PA	(MCKE/) MCKERSIE B.	
XX		
PI	Mankin L;	
XX		
DR	WPI; 2002-280939/32.	
XX		
PT	New self-excising polynucleotides, useful for producing transgenic	

filed instant US

PT plants, removing transgenes from these plants or crops (e.g. food  
PT commodities), and restricting the distribution of transgenes within the  
PT environment

XX Claim 5; Fig 7; 60pp; English.

XX The invention describes an isolated excisable polynucleotide, which  
CC comprises a desired trait polynucleotide and a recombinase polynucleotide  
CC operably linked to a promoter, all flanked by a pair of directly oriented  
CC recombination sites where the recombinase activity is regulatable. The  
CC self-excising polynucleotide is useful for producing transgenic plants,  
-CC particularly for removing all transgenic nucleic acid sequences that  
CC encode a gene product from the transgenic plant, thus restoring the  
CC original genetic configuration of the genome. The polynucleotide is also  
CC useful in methods for eliminating unwanted nucleic acids in agricultural  
CC food products and for preventing the escape of certain transgenic traits  
CC into the environment. This sequence encodes the novel phic31 integrase  
CC phic31int\*INT, created for use in the self-excising polynucleotide  
CC described in the invention.

XX Sequence 2031 BP; 454 A; 558 C; 640 G; 379 T; 0 other;

Query Match 100.0%; Score 2031; DB 24; Length 2031;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2031; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCACAAGGGGTGTGACCGGGGTGATACGTAAGTTCTCTCTTCTACCTTGATATA 60  
Db 1 ATGGCACAAGGGGTGTGACCGGGGTGATACGTAAGTTCTCTCTTCTACCTTGATATA 60  
QY 61 TATATAAT 120  
Db 61 TATATAAT 120  
QY 121 AAAGATGTAGTATATAGCAATGCTTCTCTGTAGTTTATTAAGTGTATATTTTAAATT 180  
Db 121 AAAGATGTAGTATATAGCAATGCTTCTCTGTAGTTTATTAAGTGTATATTTTAAATT 180  
QY 181 ATAACTTTTCTAATATATATGACCAAAATTTGTATGTGACAGTACGGGTGCTTACGAC 240  
Db 181 ATAACTTTTCTAATATATATGACCAAAATTTGTATGTGACAGTACGGGTGCTTACGAC 240  
QY 241 CGTCAGTCGCGCGGCGAGATAGCATGACGACGACGACGACGACGACGACGACGACGAC 300  
Db 241 CGTCAGTCGCGCGGCGAGATAGCATGACGACGACGACGACGACGACGACGACGACGAC 300  
QY 301 AACGAAGACAAGGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
Db 301 AACGAAGACAAGGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
QY 361 TTCGTGGGCAATTTAGCGAAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 420  
Db 361 TTCGTGGGCAATTTAGCGAAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 420  
QY 421 GAGTTCGAACGCATCTGACGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAT 480  
Db 421 GAGTTCGAACGCATCTGACGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAT 480  
QY 481 GACGTGTCGCGCTTCTCGCGCGCTGAAAGGTCATGACGCGGATCCGATTGTCGGAATTG 540  
Db 481 GACGTGTCGCGCTTCTCGCGCGCTGAAAGGTCATGACGCGGATCCGATTGTCGGAATTG 540  
QY 541 CTCGCGCGCGCGGCGTACGATTTGTTCCACTCAGGAAGCGCTCTTCGCGGAGAAACGTC 600  
Db 541 CTCGCGCGCGCGGCGTACGATTTGTTCCACTCAGGAAGCGCTCTTCGCGGAGAAACGTC 600  
QY 601 ATGACCTGATTCACCTGATATGCGGCTCGACGCGCTCGACACAAGAATCTTCGCTGAG 660  
Db 601 ATGACCTGATTCACCTGATATGCGGCTCGACGCGCTCGACACAAGAATCTTCGCTGAG 660  
QY 661 TCGGCGAAGATTCTGACACAGAAGACCTTCAGCGCGAATTGGCGGCTACGTCGGCGGG 720  
Db 661 TCGGCGAAGATTCTGACACAGAAGACCTTCAGCGCGAATTGGCGGCTACGTCGGCGGG 720

QY 721 AAGCGCCTTACGCTTCGAGCTTGTTCGAGACGAAAGAGATACCGCGCAACGCGCGA 780  
Db 721 AAGCGCCTTACGCTTCGAGCTTGTTCGAGACGAAAGAGATACCGCGCAACGCGCGA 780  
QY 781 ATGTCATATGTCGTCATCAACAAGTTAGCGCACTGACCACTCCCTTACCGGACCTTC 840  
Db 781 ATGTCATATGTCGTCATCAACAAGTTAGCGCACTGACCACTCCCTTACCGGACCTTC 840  
QY 841 GAGTTCGAGCGCGGCGAGTATCCGCTGTGTGCGGTGAGATGACGACGACCAACACCTT 900  
Db 841 GAGTTCGAGCGCGGCGAGTATCCGCTGTGTGCGGTGAGATGACGACGACCAACACCTT 900  
QY 901 CCCTTCAAGCGCGGCGAGTCAAGCGCGCATTCACCGGGGAGCATCACGGGCTTTGTAAG 960  
Db 901 CCCTTCAAGCGCGGCGAGTCAAGCGCGCATTCACCGGGGAGCATCACGGGCTTTGTAAG 960  
QY 961 CGCATGAGCGCTGACGCGCGGTGCGCGACCGGGGCGAGACGATTGGGAAGAAGACCGCTTCA 1020  
Db 961 CGCATGAGCGCTGACGCGCGGTGCGCGACCGGGGCGAGACGATTGGGAAGAAGACCGCTTCA 1020  
QY 1021 AGCGCTGGGACCGCGGCAACCGCTTATGCGAATCCTTCGGGACCGCGTATTGCGGCTTC 1080  
Db 1021 AGCGCTGGGACCGCGGCAACCGCTTATGCGAATCCTTCGGGACCGCGGCTATTGCGGCTTC 1080  
QY 1081 GCGGCTGAGGTGATCTACAAGAAGACCGGACGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
Db 1081 GCGGCTGAGGTGATCTACAAGAAGACCGGACGCGGCGGCGGCGGCGGCGGCGGCGG 1140  
QY 1141 TACCGCATTCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
Db 1141 TACCGCATTCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200  
QY 1201 ATGAGCGCGCTGAGTGTATGAGCTTCAGGCGGTGTTGAGCGGCGGCGGCGGCGGCAAG 1260  
Db 1201 ATGAGCGCGCTGAGTGTATGAGCTTCAGGCGGTGTTGAGCGGCGGCGGCGGCGGCAAG 1260  
QY 1261 GGGCTTTCGCGGGGGGCAAGCCATTCTGTCCGCCATGACCAAGCTGTACTGCGAGTGGC 1320  
Db 1261 GGGCTTTCGCGGGGGGCAAGCCATTCTGTCCGCCATGACCAAGCTGTACTGCGAGTGGC 1320  
QY 1321 GCGGTACATGACTTCGAAGCGCGGGGGAAGAATGATCAAGGACTTTACCGCTGCGCTGC 1380  
Db 1321 GCGGTACATGACTTCGAAGCGCGGGGGAAGAATGATCAAGGACTTTACCGCTGCGCTGC 1380  
QY 1381 CGGAAGTGTGTCGACCCGCTCCGCACTGCGGACGACGAAAGGCAAGCTGACGATG 1440  
Db 1381 CGGAAGTGTGTCGACCCGCTCCGCACTGCGGACGACGAAAGGCAAGCTGACGATG 1440  
QY 1441 GCGGCACTCGACAAAGTTGCTGCGGAAGCGCATCTTCAACAAGATCAGGCAAGCGG 1500  
Db 1441 GCGGCACTCGACAAAGTTGCTGCGGAAGCGCATCTTCAACAAGATCAGGCAAGCGG 1500  
QY 1501 GACGAAGAGACGTTGGCGCTTCTGTGGGAAGCGCGCGGCGGCGGCGGCGGCGGCGG 1560  
Db 1501 GACGAAGAGACGTTGGCGCTTCTGTGGGAAGCGCGCGGCGGCGGCGGCGGCGGCGG 1560  
QY 1561 GCGCCTGAGAAGAGCGGCGGAACGCGGCAACCTTGTGCGGAGCGCGCGGCGGCGGCGG 1620  
Db 1561 GCGCCTGAGAAGAGCGGCGGAACGCGGCAACCTTGTGCGGAGCGCGCGGCGGCGGCGG 1620  
QY 1621 GCGCTTGAAGAGCTGTACGAAGACGCGGCGGAGAGCTTACGAGGAGACCGGCTGAGG 1680  
Db 1621 GCGCTTGAAGAGCTGTACGAAGACGCGGCGGAGAGCTTACGAGGAGACCGGCTGAGG 1680  
QY 1681 AAGCACTTCCGGAAGACAAGGACGCGCTGACGCTCCGGGACGAAAGGCGGGAAGAGCGG 1740  
Db 1681 AAGCACTTCCGGAAGACAAGGACGCGCTGACGCTCCGGGACGAAAGGCGGGAAGAGCGG 1740  
QY 1741 CTTGCGGAACCTTGAAGCGCGGGAAGCGCGGCAAGTTGCCCTTGAACCAATGTTCCCGGAA 1800  
Db 1741 CTTGCGGAACCTTGAAGCGCGGGAAGCGCGGCAAGTTGCCCTTGAACCAATGTTCCCGGAA 1800

QY 1801 GACCCGACGCTGACCCGACGGCCCTAAGTCGTGTGGGGGGCGCGCTCAGTAGACGAC 1860  
|||||  
Db 1801 GACGCGACGCTGACCCGACGGCCCTAAGTCGTGTGGGGGGCGCGCTCAGTAGACGAC 1860  
QY 1861 AAGCGCGTGTTCGTGGGCTCTTCGTAGACAAGATCGTTGTACAGAACTGACTACGGGC 1920  
|||||  
Db 1861 AAGCGCGTGTTCGTGGGCTCTTCGTAGACAAGATCGTTGTACAGAACTGACTACGGGC 1920  
QY 1921 AGGGGGCAGGGAACGCCCATCGAAGAAGCGCGCTTCGATCAGTGGGCGAAGCCGCCGACC 1980  
|||||  
Db 1921 AGGGGGCAGGGAACGCCCATCGAAGAAGCGCGCTTCGATCAGTGGGCGAAGCCGCCGACC 1980  
QY 1981 GACGACGACGAAGACGACGCCCGACGGGACCGAAGACGTAGCGGCGTAG 2031  
|||||  
Db 1981 GACGACGACGAAGACGACGCCCGACGGGACCGAAGACGTAGCGGCGTAG 2031

RESULT 2

ABK12555  
ID ABK12555 standard; DNA; 2031 BP.

XX ABK12555;

DT 18-JUN-2002 (first entry)

DE DNA encoding novel phi C31 integrase, phiC31IntInt.

XX Phi C31 integrase; recombinase; transgenic; plant;

KW agricultural food product; self-excising polynucleotide;

KW gene; ds; phiC31IntInt.

OS Synthetic.

XX WO200216609-A2.

XX 28-FEB-2002.

XX 27-AUG-2001; 2001WO-US26738.

XX 25-AUG-2000; 2000US-227961P.

XX (BADI ) BASF PLANT SCI GMBH.

XX (MCKE/) MCKERSIE B.

XX Mankin L;

XX WPI; 2002-280939/32.

XX New self-excising polynucleotides, useful for producing transgenic

XX plants, removing transgenes from these plants or crops (e.g. food

XX commodities), and restricting the distribution of transgenes within the

XX environment

XX Claim 5; Fig 6; 60pp; English.

XX The invention describes an isolated excisable polynucleotide, which

XX comprises a desired trait polynucleotide and a recombinase polynucleotide

XX operably linked to a promoter, all flanked by a pair of directly oriented

XX recombination sites where the recombinase activity is regulatable. The

XX self-excising polynucleotide is useful for producing transgenic plants,

XX particularly for removing all transgenic nucleic acid sequences that

XX encode a gene product from the transgenic plant, thus restoring the

XX original genetic configuration of the genome. The polynucleotide is also

XX useful in methods for eliminating unwanted nucleic acids in agricultural

XX food products and for preventing the escape of certain transgenic traits

XX into the environment. This sequence encodes the novel phiC31 integrase

XX phiC31IntInt, created for use in the self-excising polynucleotide

XX described in the invention.

XX Sequence 2031 BP; 451 A; 562 C; 640 G; 378 T; 0 other;

XX Query Match 99.2%; Score 2015; DB 24; Length 2031;

XX Best Local Similarity 99.5%; Pred. No. 0;

Matches 2021; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ATGCACAAGGGGTTGTGACCGGGTGATACGTAAGTTCTGCTTCTACCTTGTATATA 60  
|||||  
Db 1 ATGCACAAGGGGTTGTGACCGGGTGATACGTAAGTTCTGCTTCTACCTTGTATATA 60  
QY 61 TATATAATAATTATCATTAATTAGTAGTAATATAATATTCAAAATATTTTCAAAATA 120  
|||||  
Db 61 TATATAATAATTATCATTAATTAGTAGTAATATAATATTCAAAATATTTTCAAAATA 120  
QY 121 AAAGATGTAGTATATAGCAATTCCTTTCTGTAGTTTATAAGTGTATATTTAATT 180  
|||||  
Db 121 AAAGATGTAGTATATAGCAATTCCTTTCTGTAGTTTATAAGTGTATATTTAATT 180  
QY 181 ATAACCTTTCTAATATATATGACCAAAATTTGTGATGTGACAGGTACGGGGTGTACGAC 240  
|||||  
Db 181 ATAACCTTTCTAATATATATGACCAAAATTTGTGATGTGACAGGTACGGGGTGTACGAC 240  
QY 241 CGTCAGTCGCGCGAGCGCGAGAAATAGCAGTGCAGCAAGCCCGACACAGCGTAGCGCC 300  
|||||  
Db 241 CGTCAGTCGCGCGAGCGCGAGAAATAGCAGTGCAGCAAGCCCGACACAGCGTAGCGCC 300  
QY 301 AACGAACAAGCGCGCGACCTTCAGCGGAAAGTCAGCGGAGCGGCGGCGGCTTCAGG 360  
|||||  
Db 301 AACGAACAAGCGCGCGACCTTCAGCGGAAAGTCAGCGGAGCGGCGGCGGCTTCAGG 360  
QY 361 TTCGTCGCGCATTTACGCGAAGCGCGGCGACGTCGCGCTTCGGGAGCGGCGGCGCCG 420  
|||||  
Db 361 TTCGTCGCGCATTTACGCGAAGCGCGGCGACGTCGCGCTTCGGGAGCGGCGGCGCCG 420  
QY 421 GAGTTGCAACGCATCTGAAACGAATGCCGCGCGCGCGCGCTCAACATGATCATTTCTAT 480  
|||||  
Db 421 GAGTTGCAACGCATCTGAAACGAATGCCGCGCGCGCGCTCAACATGATCATTTCTAT 480  
QY 481 GACGTGCGCGCTTCTCGCGCTGAAAGTTCATGAGCGCGCATTCGATTTCTGGAATTG 540  
|||||  
Db 481 GACGTGCGCGCTTCTCGCGCTGAAAGTTCATGAGCGCGCATTCGATTTCTGGAATTG 540  
QY 541 CTCGCCCTGGGCGTGACGATTTGTTCCACTCAGGAAGGCGTCTTCGGCAGGAAACGTC 600  
|||||  
Db 541 CTCGCCCTGGGCGTGACGATTTGTTCCACTCAGGAAGGCGTCTTCGGCAGGAAACGTC 600  
QY 601 ATGACCTGATTCACCTGATTTATCGCGCTCGACGCGCTCGCAAAAGATCTTCCTGAAG 660  
|||||  
Db 601 ATGACCTGATTCACCTGATTTATCGCGCTCGACGCGCTCGCAAAAGATCTTCCTGAAG 660  
QY 661 TCGGCGAAGATTTCTGACACGAAGAACCCTTCAGCGGCAATTGGCGGCTACGTCGGGG 720  
|||||  
Db 661 TCGGCGAAGATTTCTGACACGAAGAACCCTTCAGCGGCAATTGGCGGCTACGTCGGGG 720  
QY 721 AAGGCGCTTACGGCTTCGAGCTTTGTTGAGAGCAAGGAGATCACGCGCAAGCGGCCGA 780  
|||||  
Db 721 AAGGCGCTTACGGCTTCGAGCTTTGTTGAGAGCAAGGAGATCACGCGCAAGCGGCCGA 780  
QY 781 ATGTCATATGTCATCAACAAGTTAGCGCACTGACCACTCCCTTACCGGACCTTC 840  
|||||  
Db 781 ATGTCATATGTCATCAACAAGTTAGCGCACTGACCACTCCCTTACCGGACCTTC 840  
QY 841 GAGTTCAGCGCGACGTAATCCGCTGCTGCTGCGGTGAGATCAAGACGACAAACACTT 900  
|||||  
Db 841 GAGTTCAGCGCGACGTAATCCGCTGCTGCTGCGGTGAGATCAAGACGACAAACACTT 900  
QY 901 CCCTTCAAGCCGGGCGAGTCAAGCCGCCATTCACCCGGGCGACATCACGGGCTTTGTAAG 960  
|||||  
Db 901 CCCTTCAAGCCGGGCGAGTCAAGCCGCCATTCACCCGGGCGACATCACGGGCTTTGTAAG 960  
QY 961 CGCATGACGCTGACGCGCTGCGGACCCGGGCGAGAGACATTTGGGAAGAAGACCGCTTCA 1020  
|||||  
Db 961 CGCATGACGCTGACGCGCTGCGGACCCGGGCGAGAGACATTTGGGAAGAAGACCGCTTCA 1020  
QY 1021 AGCGCCTGGGACCCGGCAACCGTTATGCGAATCTTTCGGGACCCGGGTATTCGGGCTTC 1080  
|||||  
Db 1021 AGCGCCTGGGACCCGGCAACCGTTATGCGAATCTTTCGGGACCCGGGTATTCGGGCTTC 1080



QY 1081 GCCGCTGAGGTGATCTACAAGAAGAGCCGACGGCACGCCGACACGACATGAGGCT 1140  
DB 1081 GCCGCTGAGGTGATCTACAAGAAGAGCCGACGGCACGCCGACACGACATGAGGCT 1140  
QY 1141 TACCGCATTCAGCCGCGACCCGATCAGCTCCGCGCGTCCGAGCTTGATTCGCGACCGATC 1200  
DB 1141 TACCGCATTCAGCCGCGACCCGATCAGCTCCGCGCGTCCGAGCTTGATTCGCGACCGATC 1200  
QY 1201 ATCGAGCCCGCTGAGTGGTATGAGCTTCAGGCGGTGGACGGGCGCGGCAAG 1260  
DB 1201 ATCGAGCCCGCTGAGTGGTATGAGCTTCAGGCGGTGGACGGGCGCGGCAAG 1260  
QY 1261 GGGCTTCCCGGGGCGCAAGCCATTCCTGTCGCCCATGACAAAGCTGTACTGCGAGTGGC 1320  
DB 1261 GGGCTTCCCGGGGCGCAAGCCATTCCTGTCGCCCATGACAAAGCTGTACTGCGAGTGGC 1320  
QY 1321 GCCGTCATGACTTCGAAGCGCGGGGAGAATCGATCAAGACTCTTACCGCTGCCGTGC 1380  
DB 1321 GCCGTCATGACTTCGAAGCGCGGGGAGAATCGATCAAGACTCTTACCGCTGCCGTGC 1380  
QY 1381 CGGAAGGTGTCGACCCGCTCCGACCTGGGACGACGAAAGGACAGTGCACATG 1440  
DB 1381 CGGAAGGTGTCGACCCGCTCCGACCTGGGACGACGAAAGGACAGTGCACATG 1440  
QY 1441 GCGGCACCTCGACAAGTTCGTTGCGGAACGCACTTCAACAAGATCAGGACGCGCAAGGC 1500  
DB 1441 GCGGCACCTCGACAAGTTCGTTGCGGAACGCACTTCAACAAGATCAGGACGCGCAAGGC 1500  
QY 1501 GACGAAGAGACGTTGGCGCTTCTGTGGGAAGCCGCCGACGCTTCGGCAAGCTCAGTGA 1560  
DB 1501 GACGAAGAGACGTTGGCGCTTCTGTGGGAAGCCGCCGACGCTTCGGCAAGCTCAGTGA 1560  
QY 1561 GCGCCTGAGAAGAGCGCGGAACGCGGAACCTTGTGGGAGCGCGCGGCGGCAAGGC 1620  
DB 1561 GCGCCTGAGAAGAGCGCGGAACGCGGAACCTTGTGGGAGCGCGCGGCGGCAAGGC 1620  
QY 1621 GCCCTTGAAGAGCTGTACGAAGACCGCGCGGACGAGGCTTACGACGACCCGTTGCCAGG 1680  
DB 1621 GCCCTTGAAGAGCTGTACGAAGACCGCGCGGACGAGGCTTACGACGACCCGTTGCCAGG 1680  
QY 1681 AAGCACTTCCGGAAGCAACAGGACGCGCTGACGCTCCGCGACCAAGGGCGGAAGAGCGG 1740  
DB 1681 AAGCACTTCCGGAAGCAACAGGACGCGCTGACGCTCCGCGACCAAGGGCGGAAGAGCGG 1740  
QY 1741 CTTGCCGAACTTGAAGCCCGCGGAAGCCCGGAAGTGGCCCTTGACCAATGTTCCCGAA 1800  
DB 1741 CTTGCCGAACTTGAAGCCCGCGGAAGCCCGGAAGTGGCCCTTGACCAATGTTCCCGAA 1800  
QY 1801 GACGCCGACGCTGACCCGCGGCTTAAGTGTGGTGGGGCGCGGCTCAGTAGACGAC 1860  
DB 1801 GACGCCGACGCTGACCCGCGGCTTAAGTGTGGTGGGGCGCGGCTCAGTAGACGAC 1860  
QY 1861 AAGCGGCTGTCGTGGGCTCTTCGTAGACAAGATCGTTGTCAGGAAGTGCATACGGGC 1920  
DB 1861 AAGCGGCTGTCGTGGGCTCTTCGTAGACAAGATCGTTGTCAGGAAGTGCATACGGGC 1920  
QY 1921 AGGGGGCAGGGAAAGCCCATCGAAAGCGCGCTTCGATCAGTGGGCGAAGCCCGCGAC 1980  
DB 1921 AGGGGGCAGGGAAAGCCCATCGAAAGCGCGCTTCGATCAGTGGGCGAAGCCCGCGAC 1980  
QY 1981 GACGACGACGAAGACGACGCGCCAGGACGCGACGGAAGAGCTAGCGCGTAG 2031  
DB 1981 GACGACGACGAAGACGACGCGCCAGGACGCGACGGAAGAGCTAGCGCGTAG 2031

RESULT 3  
ABK12557  
ID ABK12557 standard; DNA; 9880 BP.  
XX  
AC ABK12557;  
XX  
DT 18-JUN-2002 (first entry)

XX DE Plasmid pBPS EW051 T-DNA region.  
XX KW Phi C31 integrase; recombinase; transgenic; plant;  
KW agricultural food product; self-excising polynucleotide;  
KW plasmid; ds; pBPS EW051; T-DNA; cyclic; circular.  
XX OS Synthetic.  
XX FH Key  
FT repeat\_unit  
FT Location/Qualifiers  
FT /tag= a  
FT /function= "The left T-DNA border and the right T-DNA  
FT border represent the end of the  
FT self-excising polynucleotide"  
FT 225..259  
FT /tag= b  
FT /label= attB  
FT /note= "Recombinase target site"  
FT 485..273  
FT /tag= c  
FT /label= g7pA  
FT 2288..519  
FT /tag= d  
FT /product= "codA-aacCI translational fusion protein"  
FT 2898..2303  
FT /tag= e  
FT /label= Nopaline\_synthase\_promoter  
FT 2925..3236  
FT /tag= f  
FT /label= Octopine\_synthase\_promoter  
FT 3260..4267  
FT /tag= g  
FT /product= "tTA"  
FT 4292..4558  
FT /tag= h  
FT /label= Nopaline\_synthase\_terminator  
FT 4597..4933  
FT /tag= i  
FT /label= Top\_10\_promoter  
FT 4977..7007  
FT /tag= j  
FT /product= "Phi\_C31\_intINT"  
FT /note= "Novel phi C31 integrase"  
FT 7027..7221  
FT /tag= k  
FT /label= Octopine\_synthase\_terminator  
FT 7253..8392  
FT /tag= l  
FT /label= Super\_Promoter  
FT 8413..9405  
FT /tag= m  
FT /product= "ergFP7INT"  
FT 9411..9677  
FT /tag= n  
FT /label= Nopaline\_synthase\_terminator  
FT 9690..9728  
FT /tag= o  
FT /label= attP  
FT /note= "Recombinase target site"  
FT 9735..9880  
FT /tag= p  
FT /function= "The right T-DNA border and the left T-DNA  
FT border represent the end of the  
FT self-excising polynucleotide"  
XX PN WO200216609-A2.  
XX XX  
XX PD 28-FEB-2002.  
XX XX  
XX PF 27-AUG-2001; 2001WO-US26738.  
XX XX  
XX PR 25-AUG-2000; 2000US-227961P.



QY 1681 AAGCACTTCCGGAACACAGGCGCTGACGCTCCGGCAGCAAGGGCGGAGAGCGG 1740  
DB 6657 AAGCACTTCCGGAACACAGGCGCTGACGCTCCGGCAGCAAGGGCGGAGAGCGG 6716  
QY 1741 CTTGCCGAACCTGAAGCCCGGAGAGCCCGGAGAGTGGCCCTTGACCAATGTTCCCGAA 1800  
DB 6717 CTTGCCGAACCTGAAGCCCGGAGAGCCCGGAGAGTGGCCCTTGACCAATGTTCCCGAA 6776  
QY 1801 GACGCCGACGCTGACCCGACCCGCTTAAGTCGTGGTGGGGCGCGCTCAGTAGACGAC 1860  
DB 6777 GACGCCGACGCTGACCCGACCCGCTTAAGTCGTGGTGGGGCGCGCTCAGTAGACGAC 6836  
QY 1861 AAGCGCGTGTCTGCTGGGCTCTTCTGTAACAAGATCGTTGTACGAAAGTGCAGTACGGGC 1920  
DB 6837 AAGCGCGTGTCTGCTGGGCTCTTCTGTAACAAGATCGTTGTACGAAAGTGCAGTACGGGC 6896  
QY 1921 AGGGGGCAGGGAACGCCCATCGAGAAGCGCGCTTCGATCAGTGGGGGAGCGCCGAC 1980  
DB 6897 AGGGGGCAGGGAACGCCCATCGAGAAGCGCGCTTCGATCAGTGGGGGAGCGCCGAC 6956  
QY 1981 GACGACGACGAGACGACGCGCCGACGAGCGGCGGAGAGAGTACGGCGTAG 2031  
DB 6957 GACGACGACGAGACGACGCGCCGACGAGCGGCGGAGAGAGTACGGCGTAG 7007

RESULT 4  
AAD04935

ID AAD04935 standard; DNA; 5711 BP.

XX AC AAD04935;

DT 17-JUL-2001 (first entry)

XX DE C31-Int expression vector PRK65 for phic31 integrase.

KW Gene trapping construct; conditional mutation; unidirectional inversion;  
KW recombinase recognition sequence; RRS; disruption cassette;  
KW selection cassette; transgenic organism; expression vector PRK65;  
KW phic31 integrase; C31-Int mediated inversion; ds.

OS Chimeric - Bacteriophage phi-C31.  
OS Chimeric - Cytomegalovirus.  
OS Chimeric - Unidentified.

XX FH Key Location/Qualifiers  
FT promoter 1..700

FT /tag- a  
FT /note- "Cytomegalovirus immediated early gene promoter"  
FT Intron 700..970

FT /tag- b  
FT /note- "Hybrid Intron"  
FT CDS 978..2819

FT /tag- c  
FT /product- "C31-Int, a phic31 phage derived integrase"  
FT misc\_feature 2831..3020

FT /tag- d  
FT /note- "Polyadenylation sequence"

XX PN WO200129208-A1.

XX PD 26-APR-2001.

XX PF 16-OCT-2000; 2000WO-EP10162.

XX PR 16-OCT-1999; 99EP-0120592.  
XX PR 27-OCT-1999; 99US-0162016.

XX PA (ARTE-) ARTEMIS PHARM GMBH.  
XX PA (FRAN-) FRANKGEN BIOTECHNOLOGIE AG.

XX PI Kuehn R, Von Melchener H, Altschmid J;  
XX

DR WPI; 2001-308486/32.  
XX  
PT New gene trapping construct capable of causing conditional mutations in  
PT genes, comprises functional DNA segment inserted in sense or antisense  
PT direction relative to gene to be trapped  
XX  
PS Example 2; Page 61-63; 78pp; English.  
XX

CC The present invention relates to a conditional gene trapping construct  
CC capable of causing conditional mutations in genes. The gene trapping  
CC construct comprises two functional DNA segments, each being flanked by  
CC two recombinase recognition sequences (RRSs) specific to site specific  
CC recombinase which is capable of unidirectional inversion of double  
CC standard DNA segment. One of the DNA segment (disruption cassette) is  
CC inserted in antisense orientation relative to the transcriptional  
CC orientation of the gene to be trapped. The other DNA segment (selection  
CC cassette) is inserted in sense direction relative to the transcriptional  
CC orientation of the gene to be trapped. The cell comprising the gene  
CC trapping construct is useful for the identification and/or isolation of  
CC genes. The transgenic organism comprising the gene trapping construct is  
CC useful to study gene function at various developmental stages. The gene  
CC trapping construct is useful for mutationally inactivating all cellular  
CC genes. The present sequence is an expression vector PRK65. The PRK65  
CC vector is transfected into a murine cell along with the test vector  
CC PRK73. The PRK65 is used for the expression of phic31 integrase to  
CC detect C31-Int (a phic31 phage derived integrase) mediated inversion.  
XX  
SQ Sequence 5711 BP; 1329 A; 1512 C; 1616 G; 1254 T; 0 other;

Query Match 88.5%; Score 1796.6; DB 22; Length 5711;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 1805; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 213 GATGTGACGATACGCGGGGTGCTTACGACCGCTCAGTCGCGGAGCGCGAGATAGCACTGC 272  
DB 1001 GGTGACACGATACGCGGGGTGCTTACGACCGCTCAGTCGCGGAGCGCGAGATAGCACTGC 1060  
QY 273 AGCAAGCCGACGACACAGCGGTAGCGCCCAAGAGACAAAGCGCGCGACCTTCAGCGCGA 332  
DB 1061 AGCAAGCCGACGACACAGCGGTAGCGCCCAAGAGACAAAGCGCGCGACCTTCAGCGCGA 1120  
QY 333 AGTCGAGCGGACGCGGGCGGCTTCAGGTTCTCGGGCATTTACGGAAGCGCGGCAC 392  
DB 1121 AGTCGAGCGGACGCGGGCGGCTTCAGGTTCTCGGGCATTTACGGAAGCGCGGCAC 1180  
QY 393 GTCGCGTTTCGGGACGCGCGAGCGCCCGGAGTTCGAACGCATCTGAACGAATGCCGCGC 452  
DB 1181 GTCGCGTTTCGGGACGCGCGAGCGCCCGGAGTTCGAACGCATCTGAACGAATGCCGCGC 1240  
QY 453 CGGGCGGCTCAACATGATCATTTGTCTATGACGTGTCCGCTTCTCGCGCTGAGGTCTAT 512  
DB 1241 CGGGCGGCTCAACATGATCATTTGTCTATGACGTGTCCGCTTCTCGCGCTGAGGTCTAT 1300  
QY 513 GGACGCGATTCGATTTGTCTCGGAATTTGCTCGCCCTGGCGGTGACGATTTGTTCCACTCA 572  
DB 1301 GGACGCGATTCGATTTGTCTCGGAATTTGCTCGCCCTGGCGGTGACGATTTGTTCCACTCA 1360  
QY 573 GGAAGCGCTTTCGGGACGGAAGAGCTGATGACCTGATTCACCTGATTAATGCGGCTCGA 632  
DB 1361 GGAAGCGCTTTCGGGACGGAAGAGCTGATGACCTGATTCACCTGATTAATGCGGCTCGA 1420  
QY 633 CGCGTCGACAAAGATCTTCTGTAAGTTCGGGAGATTTCTGACACGAAACCTTCA 692  
DB 1421 CGCGTCGACAAAGATCTTCTGTAAGTTCGGGAGATTTCTGACACGAAACCTTCA 1480  
QY 693 GCGGAAATTTGGGCGGTACGTCGGCGGAAGGCGCTTACGGCTTCGAGCTTGTTCGGA 752  
DB 1481 GCGGAAATTTGGGCGGTACGTCGGCGGAAGGCGCTTACGGCTTGTTCGGA 1540  
QY 753 GACGAAGAGATCAGCGGCAAGCGCGGATGTTCAATGTCATCATCAACAGTTAGCGCA 812  
DB 1541 GACGAAGAGATCAGCGGCAAGCGCGGATGTTCAATGTCATCATCAACAGTTAGCGCA 1600

QY 813 CTCGACCACTCCCTTACCGGACCCTTCGAGTTCCAGACCCCGACGTAATCCGGTGTG 872  
Db 1601 CTCGACCACTCCCTTACCGGACCCTTCGAGTTCCAGACCCCGACGTAATCCGGTGTG 1660  
QY 873 GCGTGAGATCAAGACGCAACAACACCTTCCCTTCAAGCCGGGAGTCAAGCCCATTTCA 932  
Db 1661 GCGTGAGATCAAGACGCAACAACACCTTCCCTTCAAGCCGGGAGTCAAGCCCATTTCA 1720  
QY 933 CCGCGGACGATCAGGGGCTTTGTAGCGCATGACGCTGACGCCGTGCCGACCCGGG 992  
Db 1721 CCGCGGACGATCAGGGGCTTTGTAGCGCATGACGCTGACGCCGTGCCGACCCGGG 1780  
QY 993 CGAGACGATTGGGAAGAACCGCTTCAAGCGCTGGGACCCGCAACCTTATGCGAAT 1052  
Db 1781 CGAGACGATTGGGAAGAACCGCTTCAAGCGCTGGGACCCGCAACCTTATGCGAAT 1840  
QY 1053 CCTTCGGGACCCGCTATTGCGGGCTTCGCCCTGAGGTGATCTACAAGAAGACCCGA 1112  
Db 1841 CCTTCGGGACCCGCTATTGCGGGCTTCGCCCTGAGGTGATCTACAAGAAGACCCGA 1900  
QY 1113 CGGCACGCCGACACGAGATTGAGGGTTACCGCATTCAGCGGACCCGATCAGCTCCG 1172  
Db 1901 CGGCACGCCGACACGAGATTGAGGGTTACCGCATTCAGCGGACCCGATCAGCTCCG 1960  
QY 1173 GCCGGTCGAGCTTGATTGCGGACCGATCATCGAGCCCGCTGAGTGTATGAGCTTCAGGC 1232  
Db 1961 GCCGGTCGAGCTTGATTGCGGACCGATCATCGAGCCCGCTGAGTGTATGAGCTTCAGGC 2020  
QY 1233 GTGGTTGGACGGCAGGGGGCGGGCAGAGGGGCTTCCCGGGGCAAGCCATTCTGTCGC 1292  
Db 2021 GTGGTTGGACGGCAGGGGGCGGGCAGAGGGGCTTCCCGGGGCAAGCCATTCTGTCGC 2080  
QY 1293 CATGACAAGCTGTACTGCGAGTGTTGGCGCCGTCATGACTTCGAAGCCGGGGAGAATC 1352  
Db 2081 CATGACAAGCTGTACTGCGAGTGTTGGCGCCGTCATGACTTCGAAGCCGGGGAGAATC 2140  
QY 1353 GATCAAGGACTCTTACCCTGCGCTGCGCGGAAGGTGTCGACCCGCTCCGACCTGGGCA 1412  
Db 2141 GATCAAGGACTCTTACCCTGCGCTGCGCGGAAGGTGTCGACCCGCTCCGACCTGGGCA 2200  
QY 1413 GCACGAAGGCACGTGCAAGCTGAGCATGGCGGCACTCGACAAGTTCGTCGGGAACGCA 1472  
Db 2201 GCACGAAGGCACGTGCAAGCTGAGCATGGCGGCACTCGACAAGTTCGTCGGGAACGCA 2260  
QY 1473 CTTCAACAAGATCAGGCAAGCCGGAAGGCGAGCAAGAGACGTTGGCGCTTGTGGGAAGC 1532  
Db 2261 CTTCAACAAGATCAGGCAAGCCGGAAGGCGAGCAAGAGACGTTGGCGCTTGTGGGAAGC 2320  
QY 1533 CGCCCGAGCGCTCGGCAAGCTCAGTGAAGCGCCCTGAGAAGAGCGGCGAAGCCCT 1592  
Db 2321 CGCCCGAGCGCTCGGCAAGCTCAGTGAAGCGCCCTGAGAAGAGCGGCGAAGCCCT 2380  
QY 1593 TGTTCGGGAGCGCGCCGACCCCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGGCG 1652  
Db 2381 TGTTCGGGAGCGCGCCGACCCCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGGCG 2440  
QY 1653 AGGAGCTTACGAGGAGCCGCTTGGCAGGAAGCACTTCGGAAGCAACAGCAGCGCTGAC 1712  
Db 2441 AGGAGCTTACGAGGAGCCGCTTGGCAGGAAGCACTTCGGAAGCAACAGCAGCGCTGAC 2500  
QY 1713 GCTCCGGCAGCAAGGGCGGAAGAGCGGCTTGGCGGAACCTTGAAGCCGCGGAAGCCCGAA 1772  
Db 2501 GCTCCGGCAGCAAGGGCGGAAGAGCGGCTTGGCGGAACCTTGAAGCCGCGGAAGCCCGAA 2560  
QY 1773 GTTGGCCCTTGACCAATGTTCCCGGAAGACGCGGACGCTGACCCGCGCCCTAAGTC 1832  
Db 2561 GCTTCCCTTGACCAATGTTCCCGGAAGACGCGGACGCTGACCCGCGCCCTAAGTC 2620  
QY 1833 GTGGTGGGGCGCGCTGAGTACGACGACAAGCGGCTGTTGTCGGGGCTCTCTGTAGACAA 1892  
Db 2621 GTGGTGGGGCGCGCTGAGTACGACGACAAGCGGCTGTTGTCGGGGCTCTCTGTAGACAA 2680  
QY 1893 GATCGTTGTACGAAGTGCAGTACGGGCGAGGGGGCAGGGAACGCCCATCGAAGAAGCGCGC 1952

Db 2681 GATCGTTGTACGAAGTGCAGTACGGGCGAGGGGCAAGGGAACGCCCATCGAAGCGCGC 2740  
QY 1953 TTCGATCAGTGGGGGGAAGCCCGGACCGAGCAGCAGCAAGACGAGCCGACGAGGCAC 2012  
Db 2741 TTCGATCAGTGGGGGGAAGCCCGGACCGAGCAGCAGCAAGACGAGCCGACGAGGCAC 2800  
QY 2013 GGAAGACGTAGCGGCGTAG 2031  
Db 2801 GGAAGACGTAGCGGCGTAG 2819  
RESULT 5  
AAH74879  
ID AAH74879 standard; DNA; 1842 BP.  
XX  
AC AAH74879;  
XX  
DT 29-OCT-2001 (first entry)  
XX  
DE Nucleotide sequence of a wildtype bacteriophage thetaC31 integrase.  
XX  
KW Recombinase; genetic engineering; monogenic disorder; ADA deficiency;  
KW cystic fibrosis; familial-hypercholesterolaemia; haemophilia;  
KW chronic granulomatous disease; Duchenne's muscular dystrophy;  
KW Fanconi's disease; anemia; sickle-cell anemia; Gaucher's disease;  
KW Hunter's syndrome; X-linked severe combined immunodeficiency; SCID;  
KW infectious disease; acquired disorder; tumour; cancer; integrase; ss.  
XX  
OS Bacteriophage thetaC31.  
XX  
PN WO200161049-A1.  
XX  
PD 23-AUG-2001.  
XX  
PF 16-FEB-2001; 2001WO-US05269.  
XX  
PR 18-FEB-2000; 2000US-0183759.  
XX  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
XX  
PI Calos MP, Scilmenti CR;  
XX  
DR WPI; 2001-522610/57.  
XX  
PT Identifying altered recombinase, involves hybridizing recombination sites in cells having specific coding sequence, transformed with altered recombinase gene, and isolating cells having product of the sequence -  
XX  
PS Disclosure; Fig 4; 101pp; English.  
XX  
CC The specification describes a method for identifying altered recombinases. The method comprises transforming cells with a first plasmid comprising two recombination sites and a coding sequence of interest, and a second plasmid encoding an altered recombinase; allowing recombination of the two recombination sites using the altered recombinase; screening and isolating transformed cells comprising the product of the sequence of interest; and identifying the altered recombinase. The altered recombinase is useful for site-specifically integrating a polynucleotide sequence of interest in a genome of a cell. It is also useful in genetic engineering of chromosomes of higher cells, and for the generation of transgenic cells, tissues, plants and animals. The altered recombinase is useful for treating monogenic disorders, e.g. ADA deficiency, cystic fibrosis, familial-hypercholesterolaemia, anemia, chronic granulomatous disease, Duchenne's muscular dystrophy, haemophilia, Fanconi's disease, sickle-cell anemia, Gaucher's disease, Hunter's syndrome and X-linked severe combined immunodeficiency (SCID), infectious diseases including viral and bacterial infections, acquired disorders including solid tumours and haematopoietic cancers such as leukaemias and lymphomas, and other cancers. The present sequence encodes a wild type bacteriophage thetaC31 integrase.  
XX



SQ Sequence 1842 BP; 387 A; 547 C; 616 G; 292 T; 0 other;

Query Match	88.4%;	Score 1795;	DB 22;	Length 1842;
Best Local Similarity	99.2%;	Pred. No, 0;		
Matches 1804; Conservative	0;	Mismatches 15;	Indels 0;	Gaps 0;

QY	213	GATGTGAGGTACGGCGGGTCTTACGACCGTCACTGCTGCGCGAGCGGAGATATGCACTGC	272
Db	24	GGTGGACACGTACGCGGGGTCTTACGACCGTCACTGCTGCGCGAGCGGAGATATGCACTGC	83
QY	273	AGCAAGCCGACGACACAGCGTAGCGCCAAAGAAAGAGCGCGGCGACCTTACGCGCA	332
Db	84	AGCAAGCCGACGACACAGCGTAGCGCCAAAGAAAGAGCGCGGCGACCTTACGCGCA	143
QY	333	AGTCGAGCGGACGGGGCGGTTCAAGTTCTGTCGGGCAATTCAGCGAAGCGCGGCGAC	392
Db	144	AGTCGAGCGGACGGGGCGGTTCAAGTTCTGTCGGGCAATTCAGCGAAGCGCGGCGAC	203
QY	393	GTCGGCGTTCGGGACGGCGGAGCGCCCGAGTTCCAAAGCATCTGAACGAATGCCGCGC	452
Db	204	GTCGGCGTTCGGGACGGCGGAGCGCCCGAGTTCCAAAGCATCTGAACGAATGCCGCGC	263
QY	453	CGGGCGCTCAACATGATTCATTTGTCTATGACGTGTCCGCTTCTCGCGCTGAAGTTCAT	512
Db	264	CGGGCGCTCAACATGATTCATTTGTCTATGACGTGTCCGCTTCTCGCGCTGAAGTTCAT	323
QY	513	GGACGCGATTCGATTTGTCTCGGAATTTGCTCGCCCTGGCGGTGACGATTTTCCACTCA	572
Db	324	GGACGCGATTCGATTTGTCTCGGAATTTGCTCGCCCTGGCGGTGACGATTTTCCACTCA	383
QY	573	GGAAGCGCTCTCCGGCAGGGAACGTCATGSACTGATTCACCTGATTTATGCGGCTCGA	632
Db	384	GGAAGCGCTCTCCGGCAGGGAACGTCATGSACTGATTCACCTGATTTATGCGGCTCGA	443
QY	633	CGCGTCGACAAAGAATCTTCGCTGAAGTTCGGCGAAGATTTCTGACACGAAGAACTTCA	692
Db	444	CGCGTCGACAAAGAATCTTCGCTGAAGTTCGGCGAAGATTTCTGACACGAAGAACTTCA	503
QY	693	GCGCGAATTTGGCGGGGTACGTCGCGCGGGAAGCGCCTTACGGCTTCGAGCTTGTTCGGA	752
Db	504	GCGCGAATTTGGCGGGGTACGTCGCGCGGGAAGCGCCTTACGGCTTCGAGCTTGTTCGGA	563
QY	753	GACGAAGGAGATCACGCGCAACGGCGGAATGTCATGTCTCATCAACAAGTTAGCGCA	812
Db	564	GACGAAGGAGATCACGCGCAACGGCGGAATGTCATGTCTCATCAACAAGTTAGCGCA	623
QY	813	CTCGACACTCCCTTACCAGGACCTTCGAGTTCGAGCCCGACGTAATCCGGTGTGTG	872
Db	624	CTCGACACTCCCTTACCAGGACCTTCGAGTTCGAGCCCGACGTAATCCGGTGTGTG	683
QY	873	GCGTGAGATCAAGACGACAAACACTTCCCTTCAAGCCGGGAGTCAAGCCGCCATTCA	932
Db	684	GCGTGAGATCAAGACGACAAACACTTCCCTTCAAGCCGGGAGTCAAGCCGCCATTCA	743
QY	933	CCCGGCGACATCACGGGGCTTTGTAAGCGCATGGACGCTGACGCCGTGCCGACCCGGG	992
Db	744	CCCGGCGACATCACGGGGCTTTGTAAGCGCATGGACGCTGACGCCGTGCCGACCCGGG	803
QY	993	CGAGACGATTGGGAAGAAGACCGCTTCAAGCGCCTGGGACCCGGCAACCGTTATGCGAAT	1052
Db	804	CGAGACGATTGGGAAGAAGACCGCTTCAAGCGCCTGGGACCCGGCAACCGTTATGCGAAT	863
QY	1053	CCTTCGGGACCCGCGTATTCGGGGCTTCGCCGCTGAGGTGATCTACAAGAAGACCCGGA	1112
Db	864	CCTTCGGGACCCGCGTATTCGGGGCTTCGCCGCTGAGGTGATCTACAAGAAGACCCGGA	923
QY	1113	CGGCACGCCGACACGAGATTTGAGGGTTACCGCATTCAGCGCGACCCGATCAGCTCCG	1172
Db	924	CGGCACGCCGACACGAGATTTGAGGGTTACCGCATTCAGCGCGACCCGATCAGCTCCG	983
QY	1173	GCCGCTGAGCTTGATTCGGGACCGGATCATCGAGCCCGCTGAGTGTATGAGCTTCAAGC	1232
Db	984	GCCGCTGAGCTTGATTCGGGACCGGATCATCGAGCCCGCTGAGTGTATGAGCTTCAAGC	1043

QY	1233	GTGGTTGGACGCGCAGGGGGCGCGCGCGCAAGGGGCTTTCCCGGGGGCAAGCCATTCTGTCCGC	1292
Db	1044	GTGGTTGAGCGGACAGGGGGCGCGCGCGCAAGGGGCTTTCCCGGGGGCAAGCCATTCTGTCCGC	1103
QY	1293	CATGACAACGCTGTACTGCGAGTGTGGCGCCGTCATGACTTCGAAGCGGGGGAAGAAATC	1352
Db	1104	CATGACAACGCTGTACTGCGAGTGTGGCGCCGTCATGACTTCGAAGCGGGGGAAGAAATC	1163
QY	1353	GATCAAGACTCTTACCAGTGCCTGCGCCCGGAAGGTGGTGAACCCGTCCGACCTGGGCA	1412
Db	1164	GATCAAGACTCTTACCAGTGCCTGCGCCCGGAAGGTGGTGAACCCGTCCGACCTGGGCA	1223
QY	1413	GCACGAAGGCACCTGCAACAGTCAACATGGCGGCACCTCGACAAGTTCTGTTGGGAACGCAT	1472
Db	1224	GCACGAAGGCACCTGCAACAGTCAACATGGCGGCACCTCGACAAGTTCTGTTGGGAACGCAT	1283
QY	1473	CTTCAACAAGATCAGGCGACGCGCGAAGGCGAGCAAGAGACGTTGGCGCTTCTGTGGGAAGC	1532
Db	1284	CTTCAACAAGATCAGGCGACGCGCGAAGGCGAGCAAGAGACGTTGGCGCTTCTGTGGGAAGC	1343
QY	1533	CGCCCGACGCTTCGGCAAGCTCACTGAGGCGCCTGAGAAGAAGCGGCGAAGCGGCGAACCCT	1592
Db	1344	CGCCCGACGCTTCGGCAAGCTCACTGAGGCGCCTGAGAAGAAGCGGCGAAGCGGCGAACCCT	1403
QY	1593	TGTTGGGAGAGCGCGCGCGACGCGCTGAACGCGCTTGAAGAAGCTGTACGAAGACCGCGCGGC	1652
Db	1404	TGTTGGGAGAGCGCGCGCGACGCGCTGAACGCGCTTGAAGAAGCTGTACGAAGACCGCGCGGC	1463
QY	1653	AGGAGCTTACGACGGAGACCCGTTGGCAGGAAGACACTTCCGGGAAGCAACAGCAGCGCTGAC	1712
Db	1464	AGGAGCTTACGACGGAGACCCGTTGGCAGGAAGACACTTCCGGGAAGCAACAGCAGCGCTGAC	1523
QY	1713	GCTCCGGCAGCAAGGGGGCGGAAGAGCGGCTTGCCGAACCTGAAGCCCGCGGAAGCCCGGAA	1772
Db	1524	GCTCCGGCAGCAAGGGGGCGGAAGAGCGGCTTGCCGAACCTGAAGCCCGCGGAAGCCCGGAA	1583
QY	1773	GTTGCCCTTGACCAATGTTCCCGGAAGACGCGCAAGCTGACCCGACCGCCCTAAGTC	1832
Db	1584	GCTTCCCTTGACCAATGTTCCCGGAAGACGCGCAAGCTGACCCGACCGCCCTAAGTC	1643
QY	1833	GTTGGTGGGGGCGCGCTCAGTAGACGACAAGCGCGTGTCTGTCGGGCTTCTGTAGACAA	1892
Db	1644	GTTGGTGGGGGCGCGCTCAGTAGACGACAAGCGCGTGTCTGTCGGGCTTCTGTAGACAA	1703
QY	1893	GATCGTTGTACGAAGTGCAGTACGCGGCAAGGGGGCAGGGAACGCCCATCGAGAAGCGCGC	1952
Db	1704	GATCGTTGTACGAAGTGCAGTACGCGGCAAGGGGGCAGGGAACGCCCATCGAGAAGCGCGC	1763
QY	1953	TTGATCAGTGTGGGCGGAAGCCGCGACCGACGACGACGAGGAAGACGACGCCAGGAGGGGCAC	2012
Db	1764	TTGATCAGTGTGGGCGGAAGCCGCGACCGACGACGACGAGGAAGACGACGCCAGGAGGGGCAC	1823
QY	2013	GGAAGACGTAGCGGCGTAG 2031	
Db	1824	GGAAGACGTAGCGGCGTAG 1842	

XX	AAQ06904
XX	ID AAQ06904 standard; DNA; 3401 BP.
XX	AC AAQ06904;
XX	DT 06-MAR-1991 (first entry)
XX	DE Sequence encoding site specific integrating function of actinomycete
DE	phage phi C31.
XX	KM Antibiotic production; isovaleryl; spiramycin; ds.
XX	OS Page phi C31.
XX	



PN EP403173-A.  
XX 19-DEC-1990.  
XX 08-JUN-1990; 90EP-0306260.  
XX 12-JUN-1989; 89US-0364959.  
XX (ELIL ) ELI LILLY & CO.  
PI Cox KL, Richardson MA, Kuhstoss SA, Schoner BE, Seno ET, Rao RN;  
DR WPI; 1990-377895/51.  
XX  
XX Use of site-specific integrating function of phage 0C31 - for  
PT increased prodn. of antibiotics and prodn. of hybrid antibiotics  
PT in actinomycetes  
XX  
XX Disclosure; Page 4-6; 46pp; English.  
PS  
XX  
XX Sequence may be incorporated into a plasmid with an antibiotic  
CC biosynthetic gene (eg. isovaleryl or spyrarmycin) and used to  
CC transform a microorganism with site specific integration.  
CC Plasmids PKC796, POJ243, PSK50 and PSK51 and actinomycetes  
CC transformed by them, are specifically claimed.  
XX  
XX  
SQ Sequence 3401 BP; 660 A; 1011 C; 1135 G; 594 T; 1 other;

Query Match 88.3%; Score 1793.4; DB 11; Length 3401;  
Best Local Similarity 99.1%; Pred. No. 0;  
Matches 1803; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 213 GATGTGCAGTACGGGGGCTTACGACCGTCACTCGCGGAGCGGAGAATAGCAGTGC 272  
Db 532 GGTGACACGTACGGGGGCTTACGACCGTCACTCGCGGAGCGGAGAATTCGAGCGC 591  
OY 273 AGCAAGCCGACGACACAGCGTAGCGGCAAGAGAGAGCGCGGCGGCGGCGGCGA 332  
Db 592 AGCAAGCCGACGACACAGCGTAGCGGCAAGAGAGAGCGCGGCGGCGGCGGCGA 651  
OY 333 AGTCAGCGGAGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 392  
Db 652 AGTCAGCGGAGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 711  
OY 393 GTCGGGCTTCGGGAGCGGGGCGGAGCGCGGAGTTCGAACGCATCTGAACGATGCCGCG 452  
Db 712 GTCGGGCTTCGGGAGCGGGGCGGAGCGCGGAGTTCGAACGCATCTGAACGATGCCGCG 771  
OY 453 CGGGCGGCTCAACATGATCATTTGTATGACGTGTGCGGCTTCTCGCGGCGGCGGCGG 512  
Db 772 CGGGCGGCTCAACATGATCATTTGTATGACGTGTGCGGCTTCTCGCGGCGGCGGCGG 831  
OY 513 GGACGCGATTCGATTTGTCTCGGAATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 572  
Db 832 GGACGCGATTCGATTTGTCTCGGAATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 891  
OY 573 GGAAGGCGTCTTCGGGAGGGAACGTCATGACCTGATTCACCTGATTATGCGGCTCGA 632  
Db 892 GGAAGGCGTCTTCGGGAGGGAACGTCATGACCTGATTCACCTGATTATGCGGCTCGA 951  
OY 633 CGCGTCGACAAAGAAATCTTCTGAAGTCGCGGAGAGATTCGACACGAAAGACCTTCA 692  
Db 952 CGCGTCGACAAAGAAATCTTCTGAAGTCGCGGAGAGATTCGACACGAAAGACCTTCA 1011  
OY 693 GCGGAATTCGGGCGGCTACGTGCGGCGGAGAGCGGCTTACGCGCTTCGAGCTTGTTCGGA 752  
Db 1012 GCGGAATTCGGGCGGCTACGTGCGGCGGAGAGCGGCTTACGCGCTTCGAGCTTGTTCGGA 1071  
OY 753 GACGAAGAGATCAAGCGGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 812  
Db 1072 GACGAAGAGATCAAGCGGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1131  
OY 813 CTCGACCACTCCCTTACCGGACCCCTTCGAGTTCGAGCCCGAGCGTAATCCGGTGTGTG 872

Db 1132 CTCGACCACTCCCTTACCGGACCCCTTCGAGTTCGAGCCCGAGCGTAATCCGGTGTGTG 1191  
OY 873 GCGTGAGATCAAGAGCGCAAAACACCTTCCCTTCAGCCCGGCGAGTCAAGCCGCCATTCA 932  
Db 1192 GCGTGAGATCAAGAGCGCAAAACACCTTCCCTTCAGCCCGGCGAGTCAAGCCGCCATTCA 1251  
OY 933 CCCGGGAGCATCAGCGGGCTTTGTAAGCGCATGAGCGCTGACGCCGCTGCCGACCCGGGG 992  
Db 1252 CCCGGGAGCATCAGCGGGCTTTGTAAGCGCATGAGCGCTGACGCCGCTGCCGACCCGGGG 1311  
OY 993 CGAGACGATTGGGAAGAGACCGCTTCAAGCGCCTGGAGCCCGGCAACCGTTATGCGAAT 1052  
Db 1312 CGAGACGATTGGGAAGAGACCGCTTCAAGCGCCTGGAGCCCGGCAACCGTTATGCGAAT 1371  
OY 1053 CCTTCGGGACCCCGGCTATTCGGGGCTTCGCGGCTGAGGTGATCTACAGAGAAGACCGGA 1112  
Db 1372 CCTTCGGGACCCCGGCTATTCGGGGCTTCGCGGCTGAGGTGATCTACAGAGAAGACCGGA 1431  
OY 1113 CGGCAAGCCGACGACGAGATTTGAGGCTTACCGCATTCAGCGCGGACCGCATCAGCTCCG 1172  
Db 1432 CGGCAAGCCGACGACGAGATTTGAGGCTTACCGCATTCAGCGCGGACCGCATCAGCTCCG 1491  
OY 1173 GCCGGTCGAGCTTGTATTCGGGAGCGGATCATCAGCCCGCTGAGTGTATGAGCTTCAGGC 1232  
Db 1492 GCCGGTCGAGCTTGTATTCGGGAGCGGATCATCAGCCCGCTGAGTGTATGAGCTTCAGGC 1551  
OY 1233 GTGGTTGAGCGGAGGCGGCGGCAAGGGGCTTCCCGGGGCAAGCCATCTGTCCGC 1292  
Db 1552 GTGGTTGAGCGGAGGCGGCGGCAAGGGGCTTCCCGGGGCAAGCCATCTGTCCGC 1611  
OY 1293 CATGACAAAGCTGTACTGCGAGTGTGGCGGCTCATGACTTCGAAGCGGCGGGAAGAATC 1352  
Db 1612 CATGACAAAGCTGTACTGCGAGTGTGGCGGCTCATGACTTCGAAGCGGCGGGAAGAATC 1671  
OY 1353 GATCAAGGACTCTTACCGCTGCGGCTGCGGCGGAGAGTGTGACCCGCTCCGACCTGGGCA 1412  
Db 1672 GATCAAGGACTCTTACCGCTGCGGCTGCGGCGGAGAGTGTGACCCGCTCCGACCTGGGCA 1731  
OY 1413 GCACGAAGGACGTCGACAGTCAAGCATGCGGCGGCACTGCAACAAGTTCGTCGGAAGCAT 1472  
Db 1732 GCACGAAGGACGTCGACAGTCAAGCATGCGGCGGCACTGCAACAAGTTCGTCGGAAGCAT 1791  
OY 1473 CTTCAACAAGATCAGGACGCGCGAAGGCGGAGAGAGAGCTTGGCGCTTCTGTGGAAGC 1532  
Db 1792 CTTCAACAAGATCAGGACGCGCGAAGGCGGAGAGAGAGCTTGGCGCTTCTGTGGAAGC 1851  
OY 1533 CGCCCGAGGCTTCGGCAAGCTCACTGAGGCGGCTGAGAAGAGCGGCGGAGCGGCAACCT 1592  
Db 1852 CGCCCGAGGCTTCGGCAAGCTCACTGAGGCGGCTGAGAAGAGCGGCGGAGCGGCAACCT 1911  
OY 1593 TGTTCGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1652  
Db 1912 TGTTCGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1971  
OY 1653 AGAGCTTACGACGAGCCGCTTGGCAGAGAGCACTTCGGAAGCAAGGAGCGGCTGAC 1712  
Db 1972 AGAGCTTACGACGAGCCGCTTGGCAGAGAGCACTTCGGAAGCAAGGAGCGGCTGAC 2031  
OY 1713 GCTCCGGGAGCAAGGGGCGGGAAGAGCGGCTTGGCGGAACCTTGAAGCCCGGGAAGCCCGGAA 1772  
Db 2032 GCTCCGGGAGCAAGGGGCGGGAAGAGCGGCTTGGCGGAACCTTGAAGCCCGGGAAGCCCGGAA 2091  
OY 1773 GTTCCCTTGAACCAATGCTTCCCGGAAGAGCGGCGGAGCGGCTGACCGGCGCTAAGTC 1832  
Db 2092 GCTTCCCTTGAACCAATGCTTCCCGGAAGAGCGGCGGAGCGGCTGACCGGCGCTAAGTC 2151  
OY 1833 GTGCTGGGGCGGCGGCTCAGTAGACGACAAAGCGGCTTCTGCTGGGCTTCTGTAAGCAA 1892  
Db 2152 GTGCTGGGGCGGCGGCTCAGTAGACGACAAAGCGGCTTCTGCTGGGCTTCTGTAAGCAA 2211  
OY 1893 GATCGTTGTACGAAGTCGACTACGGGAGGGGCGAGGAGCGGCAATCGAGAAGCGGCG 1952

Db 2212 GATCGTTCACGAGTGCAGTACGGGCGAGGGGCGAGGAGCCCATCGAGAGCGCGC 2271  
QY 1953 TTCGATCAGTGGGGGAGAGCCCGGACCGAGCAGCAGAGAGAGAGCGCCCGAGAGCGGCAC 2012  
Db 2272 TTCGATCAGTGGGGGAGAGCCCGGACCGAGCAGCAGAGAGAGAGCGCCCGAGAGCGGCAC 2331  
QY 2013 GGAAGACGTAGCGGCGTAG 2031  
Db 2332 GGAAGACGTAGCGGCGTAG 2350

RESULT 7  
AAH74880  
ID AAH74880 standard; DNA; 1842 BP.  
XX  
AC AAH74880;  
XX  
DT 29-OCT-2001 (first entry)  
XX  
DE Nucleotide sequence of altered recombinase 7C1.  
XX

KW Recombinase; genetic engineering; monogenic disorder; ADA deficiency;  
KW cystic fibrosis; familial-hypercholesterolaemia; haemophilia;  
KW chronic granulomatous disease; Duchenne's muscular dystrophy;  
KW Fanconi's disease; anemia; sickle-cell anemia; Gaucher's disease;  
KW Hunter's syndrome; X-linked severe combined immunodeficiency; SCID;  
KW infectious disease; acquired disorder; tumour; cancer; ss.  
XX  
OS Synthetic.  
OS Bacteriophage.  
XX

Key Location/Qualifiers  
FT CDS 1.1842  
FT /\*tag= a  
FT /product= "altered recombinase 7C1"  
XX

PN WO200161049-A1.  
PD 23-AUG-2001.  
XX  
PF 16-FEB-2001; 2001WO-US05269.  
XX  
PR 18-FEB-2000; 2000US-0183759.  
XX  
PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
PI Calos MP, Scilmentl CR;  
XX  
DR WPI; 2001-522610/57.  
XX  
P-PSDB; AAG63847.

Identifying altered recombinase, involves hybridizing recombination  
PT sites in cells having specific coding sequence, transformed with  
PT altered recombinase gene, and isolating cells having product of the  
PT sequence -

Disclosure; Fig 6; 101pp; English.

The specification describes a method for identifying altered  
CC recombinases. The method comprises transforming cells with a first  
CC plasmid comprising two recombination sites and a coding sequence of  
CC interest, and a second plasmid encoding an altered recombinase; allowing  
CC recombination of the two recombination sites using the altered  
CC recombinase; screening and isolating transformed cells comprising the  
CC product of the sequence of interest; and identifying the altered  
CC recombinase. The altered recombinase is useful for site-specifically  
CC integrating a polynucleotide sequence of interest in a genome of a cell.  
CC It is also useful in genetic engineering of chromosomes of higher cells,  
CC and for the generation of transgenic cells, tissues, plants and animals.  
CC The altered recombinase is useful for treating monogenic disorders,  
CC e.g. ADA deficiency, cystic fibrosis, familial-hypercholesterolaemia,  
CC anemia, chronic granulomatous disease, Duchenne's muscular dystrophy,  
CC haemophilia, Fanconi's disease, sickle-cell anemia, Gaucher's disease,

CC Hunter's syndrome and X-linked severe combined immunodeficiency (SCID),  
CC infectious diseases including viral and bacterial infections, acquired  
CC disorders including solid tumours and haematopoietic cancers such as  
CC leukaemias and lymphomas, and other cancers. The present sequence  
CC encodes an altered recombinase of the invention.

XX Sequence 1842 BP; 388 A; 549 C; 614 G; 291 T; 0 other;

Query Match 88.1%; Score 1790.2; DB 22; Length 1842;  
Best Local Similarity 99.0%; Pred. No. 0;  
Matches 1801; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 213 GATGTCAGTACCGGGTCTTACGACCGTCACTGCGGCGAGCGGAGATAGCAGTGC 272  
Db 24 GGTGACACAGTACCGGGTCTTACGACCGTCACTGCGGCGAGCGGAGATAGCAGTGC 83  
QY 273 AGCAAGCCAGCGACACAGCGGTAGCGCCAAAGAGACAAGCGCGCCGACCTTACGCGCA 332  
Db 84 AGCAAGCCAGCGACACAGCGGTAGCGCCAAAGAGACAAGCGCGCCGACCTTACGCGCA 143  
QY 333 AGTCAGCGCGAGCGGCGCGGTTTCAAGTTCGTGGGCAATTTACAGGAAGCGCGGCAC 392  
Db 144 AGTCAGCGCGAGCGGCGCGGTTTCAAGTTCGTGGGCAATTTACAGGAAGCGCGGCAC 203  
QY 393 GTGCGCGTTCGGAGCGGCGGAGCGCGCGGAGTTTGAACGATCTTGAACGATGCCGCGC 452  
Db 204 GTGCGCGTTCGGAGCGGCGGAGCGCGCGGAGTTTGAACGATCTTGAACGATGCCGCGC 263  
QY 453 CGGCGCGCTCAACATGATCATTTGTCTATGACGTGCGCGCTTTCGCGCTGAAGGTAT 512  
Db 264 CGGCGCGCTCAACATGATCATTTGTCTATGACGTGCGCGCTTTCGCGCTGAAGGTAT 323  
QY 513 GGACGCGATTCCGATTGCTCTCGGAATTCGCGCGCTTTCGCGCTGAAGGTAT 572  
Db 324 GGACGCGATTCCGATTGCTCTCGGAATTCGCGCGCTTTCGCGCTGAAGGTAT 383  
QY 573 GGAAGCGCTTTCGCGCGAGGAAACGTCATGACCTGATTCACCTGATTATGGCGTCGA 632  
Db 384 GGAAGCGCTTTCGCGCGAGGAAACGTCATGACCTGATTCACCTGATTATGGCGTCGA 443  
QY 633 CGCGTCGCACAAAGATCTTCGCTGAAGTCGCGGAAAGATTCGACACGAAAGACCTTCA 692  
Db 444 CGCGTCGCACAAAGATCTTCGCTGAAGTCGCGGAAAGATTCGACACGAAAGACCTTCA 503  
QY 693 GCGGCAATTGGCGGGGTACGTGCGGGAAGCGCGCTTACGCGCTTGTTCGGA 752  
Db 504 GCGGCAATTGGCGGGGTACGTGCGGGAAGCGCGCTTACGCGCTTGTTCGGA 563  
QY 753 GACGAAGGATCACGCGCAACGCGCAATGTCAATGTCTCATCAACAAGTTAGCGCA 812  
Db 564 GACGAAGGATCACGCGCAACGCGCAATGTCAATGTCTCATCAACAAGTTAGCGCA 623  
QY 813 CTCGACCATCCCTTACCGGACCTTCGAGTTTCGAGCCGACGTAATCCGTTGGTG 872  
Db 624 CTCGACCATCCCTTACCGGACCTTCGAGTTTCGAGCCGACGTAATCCGTTGGTG 683  
QY 873 GCGTGAGATCAAGACGACAAACACCTTCCCTTCAAGCGCGGAGTCAAGCCGCATTC 932  
Db 684 GCGTGAGATCAAGACGACAAACACCTTCCCTTCAAGCGCGGAGTCAAGCCGCATTC 743  
QY 933 CCCGGGACATCACGGGCTTTGTAAGCGCATGACGCGTACGCGGCGGACCCGGGG 992  
Db 744 CCCGGGACATCACGGGCTTTGTAAGCGCATGACGCGTACGCGGCGGACCCGGGG 803  
QY 993 CGAGACGATTGGAGAGAGACCGCTTCAAGCGCGCTGGGACCCGCAACCGTTATGCAAT 1052  
Db 804 CGAGACGATTGGAGAGAGACCGCTTCAAGCGCGCTGGGACCCGCAACCGTTATGCAAT 863  
QY 1053 CCTTCGGGACCGCGGTATTCGGGCTTCGCGCGTGAAGGTATTCACAGAGAGCCGGA 1112  
Db 864 CCTTCGGGACCGCGGTATTCGGGCTTCGCGCGTGAAGGTATTCACAGAGAGCCGGA 923  
QY 1113 CGGACGCGCGACGACGAAGATTTAGGGTTACCGCATTCAGCGGACCGCATCACGCTCG 1172

|||||  
Db 924 CGGCACGCCGACCAGCAGATGTAGGGTTACCCATTCAGCGCGACCCGATCACGCTCCG 983  
QY 1173 GCCCGTCGAGCTTGATTGCGGACCGATCATCGAAGCCCGCTGAGTGGTATGAGCTTCAGGC 1232  
Db 984 GCCGGTCGAGCTTGATTGCGGACCGATCATCGAAGCCCGCTGAGTGGTATGAGCTTCAGGC 1043  
QY 1233 GTGGTTGACGGGCGGCGGCGGCAAGGGGCTTCCCGGGGCAAGCCATTCTGTCCGC 1292  
Db 1044 GTGGTTGACGGGCGGCGGCGGCAAGGGGCTTCCCGGGGCAAGCCATTCTGTCCGC 1103  
QY 1293 CATGACAAAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAAGCGCGGGAAGATC 1352  
Db 1104 CATGACAAAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAAGCGCGGGAAGATC 1163  
QY 1353 GATCAAGGACTCTTACCGCTGCGCGTCCGGAAGGTGGTCGACCCGTCGCCACCTGGCA 1412  
Db 1164 GATCAAGGACTCTTACCGCTGCGCGTCCGGAAGGTGGTCGACCCGTCGCCACCTGGCA 1223  
QY 1413 GCACGAAGGACGTGCAACGTCAGCATGGCGGCACTCGACAAGTTCGTTGCCGAACGCAT 1472  
Db 1224 GCACGAAGGACGTGCAACGTCAGCATGGCGGCACTCGACAAGTTCGTTGCCGAACGCAT 1283  
QY 1473 CTTCACAAAGATCAGGACGCGCGGAAGGCGGACGAAGAGAGAGAGAGAGAGAGAGAG 1532  
Db 1284 CTTCACAAAGATCAGGACGCGCGGAAGGCGGACGAAGAGAGAGAGAGAGAGAGAGAG 1343  
QY 1533 CGCCCGACGCTTGGCAGAGCTCACTGAGGCGCGCTGAGAAGAGCGGCGGAACCT 1592  
Db 1344 CGCCCGACGCTTGGCAGAGCTCACTGAGGCGCGCTGAGAAGAGCGGCGGAACCT 1403  
QY 1593 TGTGCGAGCGCGCGCAGCCCTGAACGCGCTTGAAGAGCTGTACGAAGACCGCGGCGG 1652  
Db 1404 TGTGCGAGCGCGCGCAGCCCTGAACGCGCTTGAAGAGCTGTACGAAGACCGCGGCGG 1463  
QY 1653 AGGAGCTTACGACGAGACCCGTTGGCAGGAAGCACTTCCGGAAGCAACAGCGCGTGAC 1712  
Db 1464 AGGCGCGTACGACGAGACCCGTTGGCAGGAAGCACTTCCGGAAGCAACAGCGCGTGAC 1523  
QY 1713 GCTCCGGCAGCAAGGGCGGGAAGAGCGGCTTGCCTGAAGCTTGAAGCCCGCGGAAGCCCGAA 1772  
Db 1524 GCTCCGGCAGCAAGGGCGGGAAGAGCGGCTTGCCTGAAGCTTGAAGCCCGCGGAAGCCCGAA 1583  
QY 1773 GTTCCCTTGAACCAATGCTTCCCGGAAGAGCGCGGAGCGCTGACCCGCGGCGCTTAAGTC 1832  
Db 1584 GCTTCCCTTGAACCAATGCTTCCCGGAAGAGCGCGGAGCGCTGACCCGCGGCGCTTAAGTC 1643  
QY 1833 GTGGTGGGGGCGCGCTGAGTACGACGACAAGCGCGTGTCTGCGGGCTCTCTGTAGACAA 1892  
Db 1644 GTGGTGGGGGCGCGCTGAGTACGACGACAAGCGCGATGTCTGCGGGCTCTCTGTAGACAA 1703  
QY 1893 GATCGTTGTACGAAGTTCAGTACGAGGCGGCGGAGGAGCGCCCATCGAAGCGCGC 1952  
Db 1704 GATCGTTGTACGAAGTTCAGTACGAGGCGGCGGAGGAGCGCCCATCGAAGCGCGC 1763  
QY 1953 TTCGATCAGTGGGCGAAGCCGCGGACCGAGCAGCAGCAGCAAGAGAGAGAGAGAGAGAG 2012  
Db 1764 TTCGATCAGTGGGCGAAGCCGCGGACCGAGCAGCAGCAGCAAGAGAGAGAGAGAGAGAG 1823  
QY 2013 GGAAGACGTAGCGGCGTAG 2031  
Db 1824 GCAAGACGTAGCGGCGTAG 1842

RESULT 8  
AAH74881  
ID AAH74881 standard; DNA; 1839 BP.  
XX  
AC AAH74881;  
XX  
DT 29-OCT-2001 (first entry)  
XX  
DE Nucleotide sequence of altered recombinase 5C1.

XX  
KW Recombinase; genetic engineering; monogenic disorder; ADA deficiency;  
KW cystic fibrosis; familial-hypercholesterolaemia; haemophilia;  
KW chronic granulomatous disease; Duchenne's muscular dystrophy;  
KW Fanconi's disease; anemia; sickle-cell anemia; Gaucher's disease;  
KW Hunter's syndrome; X-linked severe combined immunodeficiency; SCID;  
KW infectious disease; acquired disorder; tumour; cancer; ss.  
XX  
OS Synthetic.  
OS Bacteriophage.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1839  
FT /\*tag= a  
FT /product= "altered recombinase 5C1"  
FT /note= "no termination codon given"  
PN WO200161049-A1.  
PN  
PD 23-AUG-2001.  
XX  
XX 16-FEB-2001; 2001WO-US05269.  
PF 18-FEB-2000; 2000US-0183759.  
PR (STRD ) UNIV LELAND STANFORD JUNIOR.  
PA Calos MP, Scilmenti CR;  
XX  
PI MPI; 2001-522610/57.  
DR P-PSDB: AAG63848.  
DR  
XX  
PT Identifying altered recombinase, involves hybridizing recombination  
PT sites in cells having specific coding sequence, transformed with  
PT altered recombinase gene, and isolating cells having product of the  
PT sequence -  
XX  
PS Disclosure; Fig 8; 101pp; English.  
XX  
CC The specification describes a method for identifying altered  
CC recombinases. The method comprises transforming cells with a first  
CC plasmid comprising two recombination sites and a coding sequence of  
CC interest, and a second plasmid encoding an altered recombinase; allowing  
CC recombination of the two recombination sites using the altered  
CC recombinase; screening and isolating transformed cells comprising the  
CC product of the sequence of interest; and identifying the altered  
CC recombinase. The altered recombinase is useful for site-specifically  
CC integrating a polynucleotide sequence of interest in a genome of a cell.  
CC It is also useful in genetic engineering of chromosomes of higher cells,  
CC and for the generation of transgenic cells, tissues, plants and animals.  
CC The altered recombinase is useful for treating monogenic disorders,  
CC e.g. ADA deficiency, cystic fibrosis, familial-hypercholesterolaemia,  
CC anemia, chronic granulomatous disease, sickle-cell anemia, Gaucher's disease,  
CC haemophilia, Fanconi's disease, Duchenne's muscular dystrophy,  
CC Hunter's syndrome and X-linked severe combined immunodeficiency (SCID),  
CC infectious diseases including viral and bacterial infections, acquired  
CC disorders including solid tumours and haematopoietic cancers such as  
CC leukaemias and lymphomas, and other cancers. The present sequence  
CC encodes an altered recombinase of the invention.  
XX  
SQ Sequence 1839 BP; 388 A; 547 C; 613 G; 291 T; 0 other;  
XX  
Query Match 87.9%; Score 1785.6; DB 22; Length 1839;  
Best local Similarity 99.0%; Pred. No. 0;  
Matches 1797; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
QY 213 GATGTCAGGTACGCGGCTTACGACCGTACGTCGCGCGGAGGAGATATGACGTGC 272  
Db 24 GGTGACACGTACGCGGCTTACGACCGTACGTCGCGCGGAGGAGATATGACGTGC 83  
QY 273 AGCAAGCCGACGACACAGCGGTAGCGCCCAAGAGAGAGGCGCGGACCTTACGCGGA 332  
Db 84 AGCAAGCCGACGACACAGCGGTAGCGCCCAAGAGAGAGGCGCGGACCTTACGCGGA 143

QY 333 AGTCGAGCGCGGAGCGGCGGCTTCAGTTCGTCGGGCATTTTCAGCGAAGCGCGGCGAC 392  
 Db 144 AGTCGAGCGCGGAGCGGCGGCTTCAGTTCGTCGGGCATTTTCAGCGAAGCGCGGCGAC 203  
 QY 393 GTCGCGCTTCGGGAGCGGCGGAGCGCGGAGTTCGAACGCATCCCTGAACGAATGCCGCGC 452  
 Db 204 GTCGCGCTTCGGGAGCGGCGGAGCGCGGAGTTCGAACGCATCCCTGAACGAATGCCGCGC 263  
 QY 453 CGGCGGCTCAACATGATCATTTGTCTATGAGCTGCGCGCTTCTCGCGCTGAAGGTAT 512  
 Db 264 CGGCGGCTCAACATGATCATTTGTCTATGAGCTGCGCGCTTCTCGCGCTGAAGGTAT 323  
 QY 513 GGAAGCGATTCGCGATGCTCGGAAATTCGCTCGCCCTGGGCGTGAAGTGTCTTCCACTCA 572  
 Db 324 GGAAGCGATTCGCGATGCTCGGAAATTCGCTCGCCCTGGGCGTGAAGTGTCTTCCACTCA 383  
 QY 573 GGAAGCGCTTCCGCGAGGGAACGCTGATGGAACCTGATTCACCTGATTTATGCGGCTCA 632  
 Db 384 GGAAGCGCTTCCGCGAGGGAACGCTGATGGAACCTGATTCACCTGATTTATGCGGCTCA 443  
 QY 633 CGCGTCGACAAAGATCTTCGCTGAAGTCGGCGAAGATTCGACACGAAGAAGCTTCA 692  
 Db 444 CGCGTCGACAAAGATCTTCGCTGAAGTCGGCGAAGATTCGACACGAAGAAGCTTCA 503  
 QY 693 GCGCGAATTCGGGCGGTACGTCGGCGGGAAGCGCGCTTACGCTTCGAGCTGTTTCGGA 752  
 Db 504 GCGCGAATTCGGGCGGTACGTCGGCGGGAAGCGCGCTTACGCTTCGAGCTGTTTCGGA 563  
 QY 753 GACGAAGGAGATCACGCGCAACGCGGAATGTCATGTCATCAACAAGTTAGCGCA 812  
 Db 564 GACGAAGGAGATCACGCGCAACGCGGAATGTCATGTCATCAACAAGTTAGCGCA 623  
 QY 813 CTCGACCACTCCCTTACCGGACCTTCGAGTTCGAGCCCGACGTAATCCGCTGTGTG 872  
 Db 624 CTCGACCACTCCCTTACCGGACCTTCGAGTTCGAGCCCGACGTAATCCGCTGTGTG 683  
 QY 873 GCGTAGATCAAGACGCAACAACTTCCCTTCAAGCCGGGAGTCAAGCCCGCATTTCA 932  
 Db 684 GCGTAGATCAAGACGCAACAACTTCCCTTCAAGCCGGGAGTCAAGCCCGCATTTCA 743  
 QY 933 CCGGCGAGCATCACGCGGCTTGTAGCGCATGACGCTGACGCGCTGCCGACCCGGGG 992  
 Db 744 CCGGCGAGCATCACGCGGCTTGTAGCGCATGACGCTGACGCGCTGCCGACCCGGGG 803  
 QY 993 CGAGACGATTTGGGAAGAAGACCGCTTCAAGCGCTGGGACCCGGCAACCGTTATGCGAAT 1052  
 Db 804 CGAGACGATTTGGGAAGAAGACCGCTTCAAGCGCTGGGACCCGGCAACCGTTATGCGAAT 863  
 QY 1053 CCTTCGGGACCCGCGTATTGCGGGCTTCGCGCTGAGGTGATCTACAAGAAGAACCCGGA 1112  
 Db 864 CCTTCGGGACCCGCGTATTGCGGGCTTCGCGCTGAGGTGATCTACAAGAAGAACCCGGA 923  
 QY 1113 CGGACGCGCGACACGGAAGATTGAGGTTACCGCATTCAGCGCGACCGCATCACGCTCCG 1172  
 Db 924 CGGACGCGCGACACGGAAGATTGAGGTTACCGCATTCAGCGCGACCGCATCACGCTCCG 983  
 QY 1173 GCGGCTCGAGCTTGATTGGGGAACCGATCAATCAAGCCCGCTGAGTGTATGAGCTTCAGGC 1232  
 Db 984 GCGGCTCGAGCTTGATTGGGGAACCGATCAATCAAGCCCGCTGAGTGTATGAGCTTCAGGC 1043  
 QY 1233 GTGTTGGACGCGGAGGCGCGGCAAGGCGCTTCCCGGGGCAAGCCATTCTGTCCGC 1292  
 Db 1044 GTGTTGGACGCGGAGGCGCGGCAAGGCGCTTCCCGGGGCAAGCCATTCTGTCCGC 1103  
 QY 1293 CATGACAAGCTGTACTCGAGTGTGCGCGCTCATGACTTCGAAGCGCGGGGAAGAATC 1352  
 Db 1104 CATGACAAGCTGTACTCGAGTGTGCGCGCTCATGACTTCGAAGCGCGGGGAAGAATC 1163  
 QY 1353 GATCAAGGACTCTTACCGCTGCCGTCGCGGGAAGGTGTCGACCCGCTCGCACCTGGGCA 1412  
 Db 1164 GATCAAGGACTCTTACCGCTGCCGTCGCGGGAAGGTGTCGACCCGCTCGCACCTGGGCA 1223

QY 1413 GCACGAAGGACGTCGACAGCTCAGCATGCGCGGACCTCGACAAAGTTCTGTCGGGAAGCAT 1472  
 Db 1224 GCACGAAGGACGTCGACAGCTCAGCATGCGCGGACCTCGACAAAGTTCTGTCGGGAAGCAT 1283  
 QY 1473 CTTCACAAGATCAGGACGCGCGGAGGCGGAGGAAGAGAGCTTGCGCTTCTGTGGGAAGC 1532  
 Db 1284 CTTCACAAGATCAGGACGCGCGGAGGCGGAGGAAGAGAGCTTGCGCTTCTGTGGGAAGC 1343  
 QY 1533 CGCCGACGCTTCGGCAAGCTTACTGAGGCGGCTGAGGAAGAGCGGGAACGCGGAACCT 1592  
 Db 1344 CGCCGACGCTTCGGCAAGCTTACTGAGGCGGCTGAGGAAGAGCGGGAACGCGGAACCT 1403  
 QY 1593 TGTTCGGAAGCGCGCGGAGCGGCTGAAAGCGCTTGAAGAGCTGTACGAAGACGCGCGGC 1652  
 Db 1404 TGTTCGGAAGCGCGCGGAGCGGCTGAAAGCGCTTGAAGAGCTGTACGAAGACGCGCGGC 1463  
 QY 1653 AGGAGCTTACGACGAGCCGCTTGGCAGGAAGCACTTCGGAAGCAACAGGACGCGTGAC 1712  
 Db 1464 AGGAGCTTACGACGAGCCGCTTGGCAGGAAGCACTTCGGAAGCAACAGGACGCGTGAC 1523  
 QY 1713 GCTCCGCGACGAAGGCGGAGAGAGCGGCTTGGCGAATTTGAAGCCGCGGAAGCCCGAA 1772  
 Db 1524 GCTCCGCGACGAAGGCGGAGAGAGCGGCTTGGCGAATTTGAAGCCGCGGAAGCCCGAA 1583  
 QY 1773 GTTCCCTTGAACCAATGTTCCCGGAGAGACGCGGACGCTGACCCGAGCCCTTAAGTC 1832  
 Db 1584 GTTCCCTTGAACCAATGTTCCCGGAGAGACGCGGACGCTGACCCGAGCCCTTAAGTC 1643  
 QY 1833 GTGTTGGGGGCGCGCTAGTAGACGACAGCGGCTGTCGCGGCTCTTCTGAGACAA 1892  
 Db 1644 GTGTTGGGGGCGCGCTAGTAGACGACAGCGGCTGTCGCGGCTCTTCTGAGACAA 1703  
 QY 1893 GATCCTTGTACGAAGTGCATACGGGACGAGGCGGAGGAAACGCGCATCGAAGCGCGC 1952  
 Db 1704 GATCCTTGTACGAAGTGCATACGGGACGAGGCGGAGGAAACGCGCATCGAAGCGCGC 1763  
 QY 1953 TTGCATCAGTGGGCGGAGCCCGGACCGGACGAGCAGCAAGAGCAGCGCCAGAGCGGAC 2012  
 Db 1764 TTGCATCAGTGGGCGGAGCCCGGACCGGACGAGCAGCAAGAGCAGCGCCAGAGCGGAC 1823  
 QY 2013 GGAAGACGTAGCGCGC 2028  
 Db 1824 GGAAGACGTAGCGCGC 1839

RESULT 9  
 AAH74882  
 ID AAH74882 standard; DNA; 1833 BP.  
 XX  
 AC AAH74882;  
 XX  
 DT 29-OCT-2001 (first entry)  
 XX  
 DE Nucleotide sequence of altered recombinase 1C1.  
 XX  
 KW Recombinase; genetic engineering; monogenic disorder; ADA deficiency;  
 KW cystic fibrosis; familial hypercholesterolaemia; haemophilia;  
 KW chronic granulomatous disease; Duchenne's muscular dystrophy;  
 KW Fanconi's disease; anemia; sickle-cell anemia; Gaucher's disease;  
 KW Hunter's syndrome; X-linked severe combined immunodeficiency; SCID;  
 KW infectious disease; acquired disorder; tumour; cancer; ss.  
 XX  
 OS Synthetic.  
 OS Bacteriophage.  
 XX  
 FH Key  
 FT CDS  
 FT  
 FT  
 XX  
 PN WO200161049-A1.  
 XX  
 PD 23-AUG-2001.

Location/Qualifiers  
 1..1833  
 /\*tag= a  
 /product= "altered recombinase 1C1"







QY 1713 GCTCCGGCAGCAAGGGCGGAGAGAGCGGCTTGCCCACTTGAAGCCGCCGAAAGCCCGAA 1772  
 Db 1524 GCTCCGGCAGCAAGGGCGGAGAGAGCGGCTTGCCCACTTGAAGCCGCCGAAAGCCCGAA 1583  
 QY 1773 GTTGGCCCTTGACCAATGGTCCCCGGAAGAGCGCCGCTGACCCGACCGCCCTAAGTC 1832  
 Db 1584 GCTTCCCTTGACCAATGGTCCCCGGAAGAGCGCCGCTGACCCGACCGCCCTAAGTC 1643  
 QY 1833 GTGGTGGGGGGCGCGCTCAGTAGACGACAAAGCCGCTGTCGGGCTCTTCGTAGACAA 1892  
 Db 1644 GTGGTGGGGGGCGCGCTCAGTAGACGACAAAGCCGCTGTCGGGCTCTTCGTAGACAA 1703  
 QY 1893 GATCGTTGTCACGAAGTCACTACGGGCGAGGGGCGAAGCGCCCATCGAAGAAGCGCGC 1952  
 Db 1704 GATGTTGTCACGAAGTCACTACGGGCGAGGGGCGAAGCGCCCATCGAAGAAGCGCGC 1763  
 QY 1953 TTCGATCAGTGGGCGAAGCGCCGACCGACGACGACGACGAAGACGACGCCCGACGCGCAC 2012  
 Db 1764 TTCGATCAGTGGGCGAAGCGCCGACCGACGACGACGACGAAGACGACGCCCGACGCGCAC 1822  
 QY 2013 GGAAGACGTAG 2023  
 Db 1823 GGAAGACGTAG 1833

RESULT 10  
 AAD19829  
 ID AAD19829 standard; DNA; 912 BP.  
 AC AAD19829;  
 DT 18-DEC-2001 (first entry)  
 DE Green fluorescent protein-1, syngFP1 gene with ST-LS1 intron sequence.  
 XX  
 KW Cestrum yellow leaf curling virus; CmYLCV; transcription;  
 KM transgenic plant; green fluorescent protein 1; GFP1; ds.  
 OS Solanum tuberosum.  
 XX  
 OS Unidentified.  
 FH Key  
 FT Intron  
 FT Location/Qualifiers  
 FT 278..465  
 FT /tag= a  
 FT /note= "Solanum tuberosum ST-LS1 intron"  
 PN WO200173087-A1.  
 -XX  
 PD 04-OCT-2001.  
 XX  
 PF 26-MAR-2001; 2001WO-EP03408.  
 XX  
 PR 27-MAR-2000; 2000GB-0007427.  
 PR 28-APR-2000; 2000GB-0010486.  
 PR 26-JAN-2001; 2001EP-0101802.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Hohn T, Stavolone L, De Haan PT, Ligon HT, Kononova M;  
 XX  
 DR WPI; 2001-616524/71.  
 XX  
 PT Novel DNA sequence obtained from genome of Cestrum yellow leaf curling  
 PT virus for conferring constitutive expression of an associated desired  
 PT polynucleotide  
 XX  
 PS Example 19; Page 66; 100pp; English.  
 XX  
 CC The invention relates to Cestrum yellow leaf curling virus (CmYLCV) novel  
 CC DNA sequences which functions as transcription promoters of an associated  
 CC polynucleotide sequence. These CmYLCV DNA molecules confers constitutive  
 CC expression of associated polynucleotide sequences. The invention also  
 CC relates to recombinant DNA sequences containing promoter sequences which

CC are used for creating transgenic plants expressing DNA of interest at all  
 CC times and in most tissues and organs. The present DNA sequence is green  
 CC fluorescent protein-1, syngFP1 gene with ST-LS1 intron sequence. plant  
 CC optimised GFP reporter gene is used in the construction of plant  
 CC transformation vectors.  
 XX  
 SQ Sequence 912 BP; 245 A; 245 C; 240 G; 182 T; 0 other;  
 Query Match 9.8%; Score 198.6; DB 22; Length 912;  
 Best Local Similarity 95.8%; Pred. No. 4.8e-35;  
 Matches 204; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
 QY 21 CGGGGTGATACGTAAGTTTCTGCTTCTTACCTTGATATATATATATATATATATATAT 80  
 Db 265 CCGGAGGGCTACGTAAGTTTCTGCTTCTTACCTTGATATATATATATATATATATATAT 324  
 QY 81 TTAGTAGTAATATATATATTTCAAAATATTTTTCAAAATAAAGAATGTAGTATATAGCA 140  
 Db 325 TTAGTAGTAATATATATATTTCAAAATATTTTTCAAAATAAAGAATGTAGTATATAGCA 384  
 QY 141 ATGCTTTTCTGAGTTTATAGTGTATATATTTTAAATTTTAACTTTCTAATATATAGCA 200  
 Db 385 ATGCTTTTCTGAGTTTATAGTGTATATATTTTAAATTTTAACTTTCTAATATATAGCA 444  
 QY 201 CCAAAATTTGTGATGTGACGTAAGCGGGTGC 233  
 Db 445 CCAAAATTTGTGATGTGACGTAAGCGGGTGC 477

RESULT 11  
 AAD19834  
 ID AAD19834 standard; DNA; 1577 BP.  
 AC AAD19834;  
 DT 18-DEC-2001 (first entry)  
 DE Promoter-reporter cassette #3 to construct plant transformation vector.  
 XX  
 KW Cestrum yellow leaf curling virus; CmYLCV; transcription;  
 KM transgenic plant; nopaline synthase; NOS; green fluorescent protein-1;  
 KW GFP1; ds.  
 XX  
 OS Chimeric - Cestrum yellow leaf curling virus.  
 OS Chimeric - Agrobacterium tumefaciens.  
 OS Chimeric - Unidentified.  
 FH Key  
 FT Promoter  
 FT Location/Qualifiers  
 FT 1..354  
 FT /tag= a  
 FT /note= "CmYLCV CmpC promoter fragment"  
 FT misc\_feature 380..1292  
 FT /tag= b  
 FT /note= "Corresponds to SyngFP1 gene"  
 FT terminator 1304..1577  
 FT /tag= c  
 FT /note= "Nopaline synthase (NOS) terminator"  
 XX  
 PN WO200173087-A1.  
 XX  
 PD 04-OCT-2001.  
 XX  
 PF 26-MAR-2001; 2001WO-EP03408.  
 XX  
 PR 27-MAR-2000; 2000GB-0007427.  
 PR 28-APR-2000; 2000GB-0010486.  
 PR 26-JAN-2001; 2001EP-0101802.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Hohn T, Stavolone L, De Haan PT, Ligon HT, Kononova M;  
 XX  
 DR WPI; 2001-616524/71.

```

XX PT Novel DNA sequence obtained from genome of Cestrum yellow leaf curling
PT virus for conferring constitutive expression of an associated desired
PT polynucleotide -
XX PS Example 19; Page 72-73; 100pp; English.
XX CC The invention relates to Cestrum yellow leaf curling virus (CmYLCV) novel
CC DNA sequences which functions as transcription promoters of an associated
CC polynucleotide sequence. These CmYLCV DNA molecules confers constitutive
CC expression of associated polynucleotide sequences. The invention also
CC relates to recombinant DNA sequences containing promoter sequences which
CC are used for creating transgenic plants expressing DNA of interest at all
CC times and in most tissues and organs. The present DNA sequence is a
CC promoter-reporter cassette which contains CmYLCV Cmpc promoter fragment,
CC green fluorescent protein-1 syngfp1 reporter gene and the
CC Agrobacterium tumefaciens nopaline synthase (NOS) terminator. Promoter-
CC reporter cassette is used to construct plant transformation vector.
XX SQ Sequence 1577 BP; 450 A; 372 C; 390 G; 365 T; 0 other;
XX
XX Query Match          9.8%; Score 198.6; DB 22; Length 1577;
XX Best Local Similarity 95.8%; Pred. No. 5.9e-35;
XX Matches 204; Conservative 0; Mismatches 9; Indels 0; Gaps 0
OY 21 CGGGGTGATACGTAAGTTTCCTGCTTCACTTGATATATATAAATTATCATTTAA 80
Db   645 CCGGAGGGCTACGTAAGTTTCCTGCTTCACTTGATATATATAAATTATCATTTAA 704
OY 81 TTAGTAGTAATATATATTTCAAATATTTTTTCAAATAAAGAATGTAGTATATAGCA 140
Db   705 TTAGTAGTAATATATATTTCAAATATTTTTTCAAATAAAGAATGTAGTATATAGCA 764
OY 141 ATTGCCTTTCTGTAGTTTATATAGTGCTATATTTTAAATTATTAACCTTTCTAATATATGA 200
Db   765 ATTCCTTTCTGTAGTTTATATAGTGCTATATTTTAAATTATTAACCTTTCTAATATATGA 824
OY 201 CCAAAATTTGTGATGTGCAGGTACGCCGGCTGC 233
Db   825 CCAAAATTTGTGATGTGCAGGTGCAGGAGCGC 857
RESULT 12
AAD19832
ID AAD19832 standard; DNA; 1618 BP.
XX AC AAD19832;
XX DT 18-DEC-2001 (first entry)
XX DE Promoter-reporter cassette #1 to construct plant transformation vector.
KW Cestrum yellow leaf curling virus; CmYLCV; transcription;
KW transgenic plant; nopaline synthase; NOS; green fluorescent protein-1;
KW GFP1; ds.
XX OS Chimeric - Cestrum yellow leaf curling virus.
OS Chimeric - Agrobacterium tumefaciens.
OS Chimeric - Unidentified.
XX FH Key Location/Qualifiers
FT Promoter 1..402
FT /tag= a
FT /note= "CmYLCV Cmps promoter fragment"
FT misc_feature 411..1331
FT /tag= b
FT /note= "Corresponds to SyngFP1 gene"
FT terminator 1343..1618
FT /tag= c
FT /note= "Nopaline synthase (NOS) terminator"
PN WO200173087-A1.
XX
```

```

PD 04-OCT-2001.
XX
XX 26-MAR-2001; 2001WO-EP03408.
XX
XX 27-MAR-2000; 2000GB-0007427.
PR 28-APR-2000; 2000GB-0010486.
PR 26-JAN-2001; 2001EP-0101802.
XX
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Hohn T, Stavolone L, De Haan PT, Ligon HT, Kononova M;
XX
XX WPI; 2001-616524/71.
XX
XX Novel DNA sequence obtained from genome of Cestrum yellow leaf curling
PT virus for conferring constitutive expression of an associated desired
PT polynucleotide
XX
XX Example 19; Page 68-69; 100pp; English.
XX
XX The invention relates to Cestrum yellow leaf curling virus (CmYLCV) novel
CC DNA sequences which functions as transcription promoters of an associated
CC polynucleotide sequence. These CmYLCV DNA molecules confers constitutive
CC expression of associated polynucleotide sequences. The invention also
CC relates to recombinant DNA sequences containing promoter sequences which
CC are used for creating transgenic plants expressing DNA of interest at all
CC times and in most tissues and organs. The present DNA sequence is a
CC promoter-reporter cassette which contains CmYLCV Cmps promoter fragment,
CC green fluorescent protein-1 syngFP1 reporter gene and the
CC Agrobacterium tumefaciens nopaline synthase (NOS) terminator. Promoter-
CC reporter cassette is used to construct plant transformation vector.
XX
XX Sequence 1618 BP; 468 A; 383 C; 400 G; 367 T; 0 other;
SQ
Query Match 9.8%; Score 198.6; DB 22; Length 1618;
Best Local Similarity 95.8%; Pred. No. 6e-35;
Matches 204; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
OY 21 CCGGGTGGATACGTAAGTTTCGCTCTACCTTGATATATATATTAATTATCATTAA 80
Db 684 CCGGAGGGCTACGTAAGTTTCGCTCTACCTTGATATATATATTAATTATCATTAA 743
OY 81 TTAGTAGTATATATATATTTTCAAAATATTTTTCAAATAAAGAATAGTATATAGCA 140
Db 744 TTAGTAGTATATATATATTTTCAAAATATTTTTCAAATAAAGAATAGTATATAGCA 803
OY 141 ATTGCTTTCTGTAGTTTATTAAGTGTATATTTTAATTATTAACCTTTCTAATATATGA 200
Db 804 ATTGCTTTCTGTAGTTTATTAAGTGTATATTTTAATTATTAACCTTTCTAATATATGA 863
OY 201 CCAAAATTTGTTGATGTGCAGGTACGGGGTGC 233
Db 864 CCAAAATTTGTTGATGTGCAGGTACGGAGCGC 896
RESULT 13
AADI19840
ID AADI19840 standard; DNA; 2943 BP.
XX
XX AADI19840;
AC
XX
XX 18-DEC-2001 (first entry)
DT
XX
XX Ubq3(At)-syngFP1-NOS terminator expression cassette.
DE
XX
XX Cestrum yellow leaf curling virus; CmYLCV; transcription; maize;
KW transgenic plant; nopaline synthase; NOS; green fluorescent protein; GFP;
KW ds.
XX
XX Chimeric - Arabidopsis thaliana.
OS
OS Chimeric - Agrobacterium tumefaciens.
OS Chimeric - Unidentified.
XX

```

FH		Key	Location/Qualifiers
FT	Promoter	1..1332	/tag= a
FT			/note- "Arabidopsis ubiquitin 3 promoter, Ubq3(At)"
FT	misc_feature	1738..2658	/tag= b
FT			/note- "Corresponds to green fluorescent protein (GFP)
FT	terminator	gene"	2670..2943
-FT		**tag= c	/note- "Nopaline synthase (NOS) terminator"
XX			
PN	WO200173087-A1.		
PD	04-OCT-2001.		
PF	26-MAR-2001; 2001WO-EP03408.		
PR	27-MAR-2000; 2000GB-0007427.		
PR	28-APR-2000; 2000GB-0010486.		
PR	26-JAN-2001; 2001EP-0101802.		
XX	(SYGN ) SYNGENTA PARTICIPATIONS AG.		
PA	Hohn T,	Stavojane L,	De Haan PT,
XX	Ligon HT,	Kononova M;	
PI	WPI; 2001-616524/71.		
DR	Noel DNA sequence obtained from genome of Cestrum yellow leaf curling virus for conferring constitutive expression of an associated desired polynucleotide -		
PT	Example 19; Page 93-94; 100pp; English.		
PS	The invention relates to Cestrum yellow leaf curling virus (CmYLCV) novel DNA sequences which functions as transcription promoters of an associated polynucleotide sequence. These CmYLCV DNA molecules confers constitutive expression of associated polynucleotide sequences. The invention also relates to recombinant DNA sequences containing promoter sequences which are used for creating transgenic plants expressing DNA of interest at all times and in most tissues and organs. The present DNA sequence is a Ubq3(At)-syngFPL-NOS terminator expression cassette which comprises Arabidopsis thaliana ubiquitin 3 promoter, Ubq3(At) operatively linked with green fluorescen tumefaciens nopaline synthase (NOS) terminator gene and Agrobacterium tumefaciens nopaline synthase (NOS) terminator. This expression cassette is used for the construction of plant transformation vectors.		
CC			
CC			
CC			
SQ	Sequence 2943 BP; 886 A; 544 C; 584 G; 929.T; 0 other;		
	Query Match	9.8%; Score 198.6; DB 22; Length 2943;	
	Best Local Similarity	95.8%; Pred. No. 7.5e-35;	
Matches	204; Conservative	0; Mismatches	9; Indels
		Gaps	0;
OY	21 CGGGGTGCATAGTAGTTTCGCTTCACTTGATATAAATAATTATCATTA	80	
Db	2011 CCGGACGGGCTACGTAAGTTCTGCTTCACTTGATATAAATAATTATCATTA	2070	
OY	81 TTAGTAGTAAATATAATTCAAATATTTTTTCAAAAATAAGAAGTAGTATATAGCA	140	
Db	2071 TTAGTAGTAAATATAATTCAAATATTTTTTCAAAAATAAGAAGTAGTATATAGCA	2130	
OY	141 ATGCTTTTCTGTAGTTTAAAGTGTATATTTTAAATTTAAACTTTCTAATATATGA	200	
Db	2131 ATTGCTTTTCTGTAGTTTAAAGTGTATATTTTAAATTTAAACTTTCTAATATATGA	2190	
OY	201 CCAAAATTTGTGATGTGCAGGTACGCCGGGTGC	233	
Db	2191 CCAAAATTTGTGATGTGCAGGTACGCCGGGTGC	2223	

```

ID AAD19830 standard; DNA; 2001 BP.
XX
AC AAD19830;
XX
DT 18-DEC-2001 (first entry)
XX
DE Beta-glucuronidase (GUS) reporter gene with Sr-Lsl intron sequence.
XX
KW Cestrum yellow leaf curling virus; CMYLCV; transcription;
KW transgenic plant; beta-glucuronidase; GUS; ds.
XX
OS Solanum tuberosum.
OS Unidentified.
XX
FH Key Location/Qualifiers
FT Intron 385..576
FT /*tag= "a
FT /note= "Solanum tuberosum Sr-Lsl intron"
XX
PN WO200173087-A1.
XX
PD 04-OCT-2001.
XX
PE 26-MAR-2001; 2001WO-EP03408.
XX
PR 27-MAR-2000; 2000GB-0007427.
PR 28-APR-2000; 2000GB-0010486.
PR 26-JAN-2001; 2001EP-0101802.
XX
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Hohn T, Stavolone L, De Haan PT, Ligón HT, Kononova M;
XX
DR WPI; 2001-616524/71.
XX
PT Novel DNA sequence obtained from genome of Cestrum yellow leaf curling
PT virus for conferring constitutive expression of an associated desired
PT polynucleotide -
XX
PS Example 19; Page 67-68; 100pp; English.
XX
CC The invention relates to Cestrum yellow leaf curling virus (CMYLCV) novel
CC DNA sequences which functions as transcription promoters of an associated
CC polynucleotide sequence. These CMYLCV DNA molecules confers constitutive
CC expression of associated polynucleotide sequences. The invention also
CC relates to recombinant DNA sequences containing promoter sequences which
CC are used for creating transgenic plants expressing DNA of interest at all
CC times and in most tissues and organs. The present DNA sequence is
CC beta-glucuronidase (GUS) reporter gene with Sr-Lsl intron sequence.
CC The GIG (GUS reporter gene with intron) gene is used in the construction
CC of plant transformation vectors.
XX
SQ Sequence 2001 BP; 532 A; 451 C; 535 G; 483 T; 0 other;

Query Match          9.6%; Score 195; DB 22; Length 2001;
Best Local Similarity 100.0%; Pred. No. 4,2e-34;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY   30 TACGTAAGTTTCGCTTCTACCTTGATATATATATAAATTATCACTTAATTAGTAGTA 89
DB   |||||||
DB   383 TACGTAAGTTTCGCTTCTACCTTGATATATATATAAATTATCACTTAATTAGTAGTA 442
QY   90 ATATAATATTTCAAATATTTTTTTTCCAATAAAAGAATGTAGTATATAGCAATTGCTTTT 149
DB   |||||||
DB   443 ATATAATATTTCAAATATTTTTTTTCCAATAAAAGAATGTAGTATATAGCAATTGCTTTT 502
QY   150 CTGTAGTTTATTAAGTGTGTATATTTTAACTTTTCTAATATATGACCAAAATTT 209
DB   |||||||
DB   503 CTGTAGTTTATTAAGTGTGTATATTTTAACTTTTCTAATATATGACCAAAATTT 562
QY   210 GTTGATGTGCAGTA 224
DB   |||||||
DB   563 GTTGATGTGCAGTA 577

```

RESULT 15  
ABK15667  
ID ABK15667 standard; DNA; 2001 BP.  
XX  
AC ABK15667;  
XX  
DT 21-MAY-2002 (first entry)  
XX  
DE Glutathione-S-transferase with S. tuberosum intron, GIG, sequence.  
XX  
KW ds; lipoxxygenase; RCI-1; transgenic; plant antifungal; pNOV2347;  
KW rice chemically induced cDNA; promoter; transit peptide; plastid;  
KW fungal mycotoxin inhibitor; plant breeding; glutathione-S-transferase;  
KW GUS; GIG.  
XX  
OS Chimeric - Escherichia coli.  
OS Chimeric - Solanum tuberosum.  
XX  
PN WO200206490-A1.  
XX  
PD 24-JAN-2002.  
XX  
PF 12-JUL-2001; 2001WO-EP08085.  
XX  
PR 13-JUL-2000; 2000GB-0017275.  
PR 15-SEP-2000; 2000GB-0022739.  
XX  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
PA (UYZU-) UNIV ZUERICH.  
XX  
PI Dudler R, Schafrath , Lawton KA;  
XX  
DR WPI; 2002-188550/24.  
XX  
PT Novel isolated nucleic acid encoding a promoter which is capable of  
PT driving chemically inducible but not wound- or pathogen-inducible  
PT expression of an associated nucleotide sequence  
XX  
PS Example 17; Page 75; 88pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (a promoter of  
CC rice chemically induced cDNA (RCI-1), which encodes a lipoxxygenase)  
CC capable of driving chemically-inducible but not wound- or pathogen-  
CC inducible expression of an associated nucleotide sequence. Also  
CC included are the RCI-1 cDNA, its encoded protein, a 4.5kb genomic clone  
CC for the lipoxxygenase gene, promoter fragments, the lipoxxygenase transit  
CC peptide which directs expressed proteins to the plastid, a vector  
CC comprising the promoter or fragments and a transgenic plant comprising  
CC the vector. The promoter or fragments are useful for expressing a  
CC nucleotide sequence of interest. The transit peptide is useful for  
CC targeting an associated protein of interest to plastids. A nucleic acid  
CC which expresses polypeptide having lipoxxygenase activity is useful for  
CC inhibiting fungal mycotoxins when transformed into a plant. The  
CC lipoxxygenase is useful for inhibiting fungal mycotoxins. The promoter is  
CC useful for regulating transcription of a chemically inducible but not  
CC wound or pathogen inducible gene, which involves applying a chemical  
CC regulator to a plant or seed containing a chemically regulatable  
CC nucleotide sequence. Transgenic plants as described above are useful for  
CC breeding improved plant lines that for example increase the effectiveness  
CC of conventional methods such as herbicide or pesticide treatment or allow  
CC to dispense with the methods due to their modified genetic properties.  
CC New crops with improved stress tolerance can be obtained that, due to  
CC their optimised genetic equipment yield harvested product of better  
CC quality than products that were not able to tolerate comparable adverse  
CC developmental conditions. The present sequence is the glutathione-S-  
CC transferase (GUS) sequence (containing an intron from solanum  
CC tuberosum, referred to as GIG) whose expression is driven from rice  
CC lipoxxygenase promoter regions in a vector construct.  
XX  
SQ Sequence 2001 BP; 532 A; 451 C; 535 G; 483 T; 0 other;

Query Match		9.6%;	Score 195;	DB 24;	length 2001;
Best Local Similarity		100.0%;	Pred. No. 4.2e-34;		
Matches 195;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	30	TACGTAAGTTTCTGCTCTACCTTTGATATATATAATAATTAATTAATTAGTAGTA	89		
Db	383	TACGTAAGTTTCTGCTCTACCTTTGATATATATAATAATTAATTAATTAGTAGTA	442		
QY	90	ATATAATATTTTCAAAATATTTTTCACAAATAAAGAATGTAGTATATAGCAATTGCTTTT	149		
Db	443	ATATAATATTTTCAAAATATTTTTCACAAATAAAGAATGTAGTATATAGCAATTGCTTTT	502		
QY	150	CTGTAGTTTATAGTGTATATTTTAAATTATAACTTTCTAATATATGACCAAAATTT	209		
Db	503	CTGTAGTTTATAGTGTATATTTTAAATTATAACTTTCTAATATATGACCAAAATTT	562		
QY	210	GTTGATGTGCAGGTA	224		
Db	563	GTTGATGTGCAGGTA	577		

Search completed: May 11, 2003, 18:05:27  
Job time : 535.5 secs





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 17:39:39 ; Search time 3307.5 Seconds  
(Without alignments)  
9944.984 Million cell updates/sec

Title: US-09-940-550A-10  
Perfect score: 2031  
Sequence: 1 atggcacaagggtgtgtgac.....cggaagacgtacgagcgtag 2031

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthum: \*  
3: em\_estlin: \*  
4: em\_estlmu: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hic: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_hic: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: em\_estfun: \*  
16: em\_estom: \*  
17: gb\_gss: \*  
18: em\_gss\_hum: \*  
19: em\_gss\_inv: \*  
20: em\_gss\_pln: \*  
21: em\_gss\_vrt: \*  
22: em\_gss\_fun: \*  
23: em\_gss\_mam: \*  
24: em\_gss\_mus: \*  
25: em\_gss\_other: \*  
26: em\_gss\_pro: \*  
27: em\_gss\_rod: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	125.8	6.2	675	13	BM109636
C 2	125.8	6.2	804	14	BQ512568
C 3	125.8	6.2	847	14	BQ512569
C 4	60.8	3.0	621	17	CNS04POX
C 5	58	2.9	928	17	CNS00DKY
C 6	58	2.9	964	17	CNS017VX

C 7	57.4	2.8	752	17	BH585665	BH585665	BOGSL20TF
C 8	57.4	2.8	986	17	CNS0170T	AL107447	Drosophila
C 9	56.8	2.8	1101	17	CNS017SY	AL108460	Drosophila
C 10	56	2.8	780	17	CNS00T74	AL089326	Arabidops
C 11	56	2.8	1201	17	CNS01038	AL098462	Drosophila
C 12	55.8	2.7	1101	17	CNS00ZAL	AL097431	Drosophila
C 13	55.6	2.7	1190	17	CNS020N7	AL206908	Tetraodon
C 14	55.4	2.7	1101	17	CNS003BD	AL064091	Drosophila
C 15	55.2	2.7	925	17	CNS0091P	AL053013	Drosophila
C 16	55.2	2.7	975	17	BH179465	BH179465	014_P_10-
C 17	55	2.7	644	17	AQ961004	AQ961004	LERF158TF
C 18	55	2.7	720	17	CNS00XZ8	AL095522	Arabidops
C 19	54.8	2.7	859	12	BF272342	BF272342	GA_EB001
C 20	54.2	2.7	1101	17	CNS00EVL	AL069706	Drosophila
C 21	54.2	2.7	1101	17	CNS00EVL	AL069706	Drosophila
C 22	54	2.7	920	17	CNS01UOL	AL168150	Tetraodon
C 23	53.8	2.6	602	17	B28108	B28108	T2G12TRD TA
C 24	53.8	2.6	1101	17	CNS00CYH	AL060100	Drosophila
C 25	53.8	2.6	1101	17	CNS002DB	AL062360	Drosophila
C 26	53.4	2.6	454	17	B96681	B96681	T3001TR TAM
C 27	53.4	2.6	748	17	AQ286687	AQ286687	RPC111-81
C 28	53.4	2.6	876	17	CNS009G1	AL053529	Drosophila
C 29	53.2	2.6	272	14	BQ596436	BQ596436	PfESTRab3
C 30	53.2	2.6	423	17	CNS00S62	AL087992	Arabidops
C 31	53.2	2.6	807	17	BH438258	BH438258	BOGLZ67TF
C 32	53.2	2.6	812	17	BH178455	BH178455	011_J-02-
C 33	53.2	2.6	812	17	CNS07KRM	AL615412	T3 end of
C 34	53.2	2.6	1101	17	CNS016LI	AL0106896	Drosophila
C 35	53	2.6	364	17	AQ583787	AQ583787	RPC1-11-4
C 36	53	2.6	1101	17	CNS00DT7	AL075293	Drosophila
C 37	52.6	2.6	1101	17	CNS00EPO	AL069493	Drosophila
C 38	52.4	2.6	673	17	CNS00UWL	AL416395	T7 end of
C 39	52.4	2.6	905	17	CNS00KHX	AL077798	Drosophila
C 40	52.2	2.6	863	12	BF264892	BF264892	HY_CEA001
C 41	52.2	2.6	1101	17	CNS00EQL	AL069526	Drosophila
C 42	52	2.6	996	17	CNS00FUH	AL071063	Drosophila
C 43	52	2.6	1043	17	CNS0145P	AL103735	Drosophila
C 44	52	2.6	1292	13	BM463105	BM463105	AGENCOURT
C 45	51.8	2.6	432	17	AQ293633	AQ293633	HS_2244_B

#### ALIGNMENTS

RESULT 1  
BM109636/c 675 bp mRNA linear EST 26-NOV-2001  
LOCUS  
DEFINITION EST57172 potato roots solanum tuberosum cDNA clone cPRO4H6 5' end,  
mRNA sequence.

ACCESSION BM109636  
VERSION BM109636.1 GI:17070714  
KEYWORDS  
SOURCE  
ORGANISM

Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE  
AUTHORS van der Hoeven, R., Sun, H., Karamycheva, S.A., Tsai, J., Van Aken, S.,  
Utterback, T., Chlemingo, A., Bougri, O., Buell, C.R., Ronning, C.,  
Tanksley, S. and Baker, B.  
TITLE Generation of ESTs from potato roots  
JOURNAL Unpublished (2001)  
COMMENT Contact: Research Genetics, Libraries Division  
Tel: 1-800-711-6195  
Email: cdna@resgen.com

FEATURES  
source  
Location/Qualifiers  
1..675  
/organism="Solanum tuberosum"  
/cultivar="Kennebec"

/db\_xref="taxon:4113"  
/clone="CPR04H6"  
/clone\_lib="potato roots"  
/tissue\_type="roots"  
/dev\_stage="in vitro grown stem cuttings"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Cornell University, Tanksley lab;  
sequencing: The Institute for Genomic Research. Roots were  
isolated from in vitro grown stem cuttings on CM medium.  
Roots were isolated two weeks after placing the stem  
cuttings from in vitro grown plants on medium."

BASE COUNT 252 a 117 c 100 g 206 t  
ORIGIN

Query Match 6.2%; Score 125.8; DB 13; Length 675;  
Best Local Similarity 83.9%; Pred. No. 3.3e-18;  
Matches 167; Conservative 0; Mismatches 27; Indels 5; Gaps 2;

QY 31 ACGTAAGTTCTGCTTCTACCTTGATATATATATAATTAATTCATTAGTAGTAA 90  
DB 663 AGGTTGTGTTACTTCTACCTTGATATATAT--AATAATATCATTAATTAGTAGTAA 607  
QY 91 TATAATATTTCAATATTTTTCAAAATAAAGAATGTAGATATAGCAATGCTTTTC 150  
DB 606 AATAATATTTCCAT--TTTTTTCAAAATAAAGAATGTAAATTAAGCAATGCTTTTC 549  
QY 151 TGTAGTTTAAAGTGTATATTTTATTAATTTTCTAATATATATGACCAAAATTG 210  
DB 548 TGTAGTTTAAAGTGTATATTTTATTAATTTTCTAATATATATGACCAAAACATG 489  
QY 211 TTGATGTGACGATACGGG 229  
DB 488 GTGATGTTTAGGAAAGG 470

RESULT 2 804 bp mRNA linear EST 10-JUN-2002  
BQ512568/c LOCUS  
DEFINITION EST619983 Generation of a set of potato CDNA clones for microarray  
analyses mixed potato tissues Solanum tuberosum CDNA clone STMH222  
5' end, mRNA sequence.  
ACCESSION BQ512568  
VERSION BQ512568.1 GI:21371437  
KEYWORDS EST.  
SOURCE potato.  
ORGANISM Solanum tuberosum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
AUTHORS Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,  
Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and  
Karameycheva,S.A.

TITLE Generation of a set of potato CDNA clones for microarray analyses  
JOURNAL Unpublished (2002)  
COMMENT Other\_ESTS: EST619984  
Contact: Robin Buell  
The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: potato@tigr.org  
This clone is available through the Research Genetics, contact the  
Research Genetics for further information 1-800-711-6195 or  
cdna@resgen.com  
Seq primer: T3.

FEATURES  
Source Location/Qualifiers  
1..804

/organism="Solanum tuberosum"  
/cultivar="Kennebec or Binjite"  
/db\_xref="taxon:4113"  
/clone="STMH222"  
/clone\_lib="Generation of a set of potato CDNA clones for  
microarray analyses mixed potato tissues"

/tissue\_type="mixed tissues"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Combination of untreated and Phytophthora  
infestans-treated libraries of stolons, leaves, leaflets,  
axillary buds of stem explants, petioles, germinating eyes  
, tubers, or roots."

BASE COUNT 284 a 149 c 116 g 255 t  
ORIGIN

Query Match 6.2%; Score 125.8; DB 14; Length 804;  
Best Local Similarity 83.9%; Pred. No. 3.4e-18;  
Matches 167; Conservative 0; Mismatches 27; Indels 5; Gaps 2;

QY 31 ACGTAAGTTCTGCTTCTACCTTGATATATATAATTAATTCATTAGTAGTAA 90  
DB 663 AGGTTGTGTTACTTCTACCTTGATATATAT--AATAATATCATTAATTAGTAGTAA 607  
QY 91 TATAATATTTCAATATTTTTCAAAATAAAGAATGTAGATATAGCAATGCTTTTC 150  
DB 606 AATAATATTTCCAT--TTTTTTCAAAATAAAGAATGTAAATTAAGCAATGCTTTTC 549  
QY 151 TGTAGTTTAAAGTGTATATTTTATTAATTTTCTAATATATGACCAAAATTG 210  
DB 548 TGTAGTTTAAAGTGTATATTTTATTAATTTTCTAATATATGACCAAAACATG 489  
QY 211 TTGATGTGACGATACGGG 229  
DB 488 GTGATGTTTAGGAAAGG 470

RESULT 3 847 bp mRNA linear EST 10-JUN-2002  
BQ512569 LOCUS  
DEFINITION EST619984 Generation of a set of potato CDNA clones for microarray  
analyses mixed potato tissues Solanum tuberosum CDNA clone STMH222  
3' end, mRNA sequence.  
ACCESSION BQ512569  
VERSION BQ512569.1 GI:21371438  
KEYWORDS EST.  
SOURCE potato.  
ORGANISM Solanum tuberosum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
AUTHORS Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,  
Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and  
Karameycheva,S.A.

TITLE Generation of a set of potato CDNA clones for microarray analyses  
JOURNAL Unpublished (2002)  
COMMENT Other\_ESTS: EST619983  
Contact: Robin Buell  
The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: potato@tigr.org  
This clone is available through the Research Genetics, contact the  
Research Genetics for further information 1-800-711-6195 or  
cdna@resgen.com  
Seq primer: T7.

FEATURES  
Source Location/Qualifiers  
1..847

/organism="Solanum tuberosum"  
/cultivar="Kennebec or Binjite"  
/db\_xref="taxon:4113"  
/clone="STMH222"  
/clone\_lib="Generation of a set of potato CDNA clones for  
microarray analyses mixed potato tissues"  
/tissue\_type="mixed tissues"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Combination of untreated and Phytophthora  
infestans-treated libraries of stolons, leaves, leaflets,



	BASE COUNT	ORIGIN
/clone_11b-"RPCI-98"	262 a	70 c 84 g 321 t 191 others
/note="end : T7"		

Query Match	2.9%;	Score 58;	DB 17;	length 928;
Best Local Similarity	36.1%;	Pred. No. 0.012;		
Matches 65;	Conservative 50;	Mismatches 65;	Indels 0;	Gaps 0;

```

0Y      38  TTTCGCTCTACCTTGGATATATATATTAATTAATGCTAATTAAGTAGATAATA  97
      ||| : || | ||| ::::::::::: | ::::| ::::::::::: | : ::||
Db      807  TTTTWTCTTTTCTTTTCTTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  748

```

QY	98	TTTCAAAATATTTTTTTTCCAAATATAAGAATGTAGTATATAGCAATTGCTTTCTGTAGTT	157
		: : :       :   :   :                 :   :	
Db	747	TWWWWWWATATTTTWTATWTWAWATAWAAAAAATAAATTTTWTWTATTTTAAWTTW	688

**Qy**    158 TATAAGTGTGATATTTTAATTATTAACCTTTCTAATATATGACCACAAATTTTGTCATGT 217  
       | | | | | : | | : | | : | | : | | :  
**Db**    687 TAAAATTTTTTAATTTTWTTTTWTATATAAWMTAWMAWAAMWATATTATAT 628  
       | | | | | : | | : | | : | | : | | :

RESULT 6	
CNS017VX/c	
LOCUS	
CNS017VX	964 bp
	DNA
	linear
	CCC 36-TTT-10000

BACN37015 of Drosophila library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.  
AL108567

KEYWORDS	GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr))

collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - . This *Drosophila melanogaster* BAC library (Dros BAC) was made by Alvin Billand at the Centre for

project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

```
source
1. .964
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
```

```

/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : SP6"

```

ORIGIN	Query Match
	2.98; Score 58; DB 17; Length 964.

Matches	76;	Conservative	29;	Mismatches	59;	Indels	0;	Gaps	0;
52	TTTGATATATATATAATTAATTATCATTAATTAGTAGTAATATATAATTTCCAATATTTTTT	111							

b6  
927 TTTTCTTTTTTTTTTTTTTTTTTWTTAATTAATTACTCMTMMATAWTTAAWTITTTTT 868

OY 112 TTCAAAATAAAGCATGTAGTATATACCAATTGCTTTTCCTAGTTTATAAGTGCGTATA 171

08 / WWWWAAAAAAAAAAATJAWATWTATAMWAAAAATTTTTTATAAATWTATTATTATAATWTA 808

QY 172 TTTTAATTATAACTTTCATATATATGACCAAAATTGGTGAT 215  
| : | :: | : || | | | : || | | |  
Db 807 TWTATAAWWWTTWACATNAAMATTACAAGAAAATTATTTAT 764

RESULT 7  
BH585665  
LOCUS  
  
PUGCCCCC  
  
ZCC L  
  
GAA  
  
T T C C A G C C C  
C C C C C C C C C

DEFINITION	BOGSL20TF BOGS Brassica oleracea genomic clone BOGSL20, DNA sequence.
ACCESSION	BH585665
VERSION	BH585665.1 GI:17838123

SOURCE ORGANISM	ISS.	1955
Brassica oleracea.		
Brassica oleracea		
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		

REFERENCE  
TOWN, C.D., VAN AKEN, S., UTTERBACK, T. and FRASER, C.M.  
1 (bases 1 to 752)  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

JOURNAL unpublished (2001)  
COMMENT other\_gsss: BOGSL20TR  
Contact: Chris Town

9/12 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

FEATURES

Location	Qualiflor
Class	sheared ends.
Seq primer	TF
UMA	is from a doublet naploid provided by Tom Osborn.

```

21: 172
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"

```

```

/clone.lib="BOGS"
/notes="Vector: site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

```

ORIGIN	Score	DB	Length
Query Match	2.88;	Score 57.4;	DB 17; Length 752;

Matches	106;	Conservative	0;	Mismatches	81;	Indels	0;	Gaps	0;
QY	34	TAAGTTTCGCTTCTACCTTGATATATAATTAATAATTATCATTTACTTAATAT	93						

DD 554 TAAATATATAAAAATAGTCTTAATTAACCAAAAAATAGTTTTAATTAATAAAAAATAGTTTA 613  
 QY 94 AATATTTCAATATATTTTTTTCAAAAATAAAGAATGTAGTATATAGCAATTCCTTTCTGT 153

20 014 AAAAAACAAAAAAGCTTTTAAATAAATATAAAAAATGGTTATTAATATACAAATTTTAT 673  
 QY 154 AGTTAAAGTGTGTATATTTTAATTTAACTTTCTAATATATGACCAAAATTTGTTG 213

074 AAAAAAAAAAAAAAAAAAAAAAATAGTTT /33

QY 214 ATGTCA 220

## RESULT 8

LOCUS	986 bp	DNA	linear	GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN17H15 of DrosBAC library from Drosophila melanogaster (fruit+			

ACCESSION	AL107447
VERSION	AL107447.1
	GI:5627417





BASE COUNT	ORIGIN	Query Match	Best Local	Matches	Score	DB	Length	Indels	Gaps
187 a	158 c	150 g	394 t	312 others	2.8%;	56;	1201;	0;	0;
<p>Query Match</p> <p>Best Local Similarity 37.2%; Pred. No. 0.036;</p> <p>Matches 74; Conservative 49; Mismatches 76; Indels 0; Gaps 0;</p>									
Db	8	AAGGGTGTGACCGGGTGATACGTAAGTTCTGCTTACCTTGATATATATAA	67						
Db	940	ATGRGKGTKTCKDKDGTTKGATTKGTGCGATWAAAAMWDKDKKKKAMRKDAKWRWAA	999						
QY	68	TAATTATCATTAATTAGTAGTAATATATTTCAATATTTTTCAAAATAAAGAAT	127						
Db	1000	KGMAWMAAANAAGAAAAATWMTATTAATDRDWATTTTTTTTTTATTTTWTWTAAAW	1059						
QY	128	GTAGTATATAGCAATTCCTTTCTGAGTTATTAAGTGTGTATATTTTAAATTAACCT	187						
Db	1060	ATWMTATATADAAAAMATATWKAAMTWMTAAATANTWTTTWTWTWTWTAAATTTT	1119						
QY	188	TTCTAATATATGACCAAAA	206						
Db	1120	TTTTTAAATATATAAAAAA	1138						
RESULT 12	CNS00ZAL	CNS00ZAL	1101 bp	DNA	linear	GSS	26-JUL-1999		
LOCUS	DEFINITION	Drosophila melanogaster genome survey, sequence SP6 end of BAC BACN01120 of Drosophila library from Drosophila melanogaster (fruit fly), genomic survey sequence.							
ACCESSION	VERSION	AL097431	GI:5609042						
KEYWORDS	SOURCE	GSS.							
ORGANISM	Drosophila melanogaster.								
	Drosophila melanogaster								
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.								
REFERENCE	1 (bases 1 to 1101)								
AUTHORS	TITLE	Genoscope.							
JOURNAL	Direct Submission								
	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)								
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros. BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.								
FEATURES	source	Location/Qualifiers							
		1. 1101							
		/organism="Drosophila melanogaster"							
		/db_xref="taxon:7227"							
		/clone="BACN01120"							
		/clone_lib="DrosBAC"							
		/plasmid="pBelobAC11"							
		/note="end : SP6"							
BASE COUNT	391 a	216 c	197 g	149 t	148 others				
ORIGIN									
Query Match	2.7%;	Score 55.8;	DB 17;	Length 1101;					
Best Local Similarity	44.3%;	Pred. No. 0.039;							
Matches	74;	Conservative 32;	Mismatches 61;	Indels 0;	Gaps 0;				

Db	914	AMWTTTTTTTTTTTTTTTWTWTVTTTTATATWTAFTWWMTKTTTTTTWAHHMAATWT	973
QY	89	AATATAATATTTCAAATATTTTTTCACAAATAAAGACTAGTATATAGCAATTGCTTT	148
Db	974	TTKWAANAANAAAAAANAAMAATAWAWATRTTTTTTTWWMWMAATATWT	1033
QY	149	TCTGTAGTTTATAGTGCTATATTTTAACTTTCTTAATA	195
Db	1034	TTTTTTTTTWTAAWTTTWTWTTTTTTTTTTTAAAWA	1080
<hr/>			
RESULT 13			
CNSO2ON7/C	Locus	CNSO2ON7	1190 bp DNA linear GSS 14-MAY-2000
DEFINITION	definition	Tetraodon nigroviridis genome survey sequence PUC-Or1 end of clone 153P04 of library G from Tetraodon nigroviridis, genomic survey sequence.	
ACCESSION	AL206908	GI:7865727	
VERSION	AL206908.1	G1:7865727	
KEYWORDS	GSS; genome survey sequence.		
SOURCE	Tetraodon nigroviridis.		
ORGANISM	Tetraodon nigroviridis		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;			
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;			
Tetraodontidae; Tetraodon.			
REFERENCE	1 (bases 1 to 1190)		
AUTHORS	Roeest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetlier,F., Saurin,W. and Weissenbach,J.		
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1190)		
AUTHORS	Roeest-Crollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetlier,F., Saurin,W., Bernot,A. and Weissenbach,J.		
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 1190)		
AUTHORS	Genoscope.		
JOURNAL	Direct Submission		
COMMENT	Submitted (12-APR-2000)  This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/Tetraodon">http://www.genoscope.cns.fr/Tetraodon</a> .		
FEATURES	Location/Qualifiers		
source	1..1190		
	/organism="Tetraodon nigroviridis"		
	/db_xref="taxon:99883"		
	/clone="153P04"		
	/clone_1b="G"		
	/note="Genoscope sequence ID : C0AG153DH02SP1-end ; PUC-Or1"		
BASE COUNT	373 a 229 c 141 g 312 t	135 others	
ORIGIN			
Query Match	2.7%; Score 55.6; DB 17; Length 1190;		
Best Local Similarity	48.8%; Pred. No. 0.044;		
Matches	82; Conservative 18; Mismatches 68; Indels 0; Gaps 0;		
QY	48	TACCTTGATATATATATAATTATCATTAATTAGTAGTAATATAATATTTCAAATAT	107
Db	824	TAMWWWATATATTATAAANANANAWMTAWMTANTATTATATATATAMWWWTTWANAAAT	765
QY	108	TTTTTTCAAATAAAGAATAGTATATATAGCAATTCCTTTCTGTAGTTTATAAGTGTG	167
Db	764	ATATATNAAAAAAAAAATAATATATATTTATATATATATATNTATWATATAATTTATA	705
QY	168	TATATTTTAATTATTAACITTTCTAATATATAGACCAAAATTTGTTGAT	215

[illegible]

VERSION AL053013.1 GI:4934461  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 925)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aaron Mammoler in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source 1..925  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="BACR19D16"  
/clone\_lib="RPCI-98"  
/note="end : TET3"

BASE COUNT 120 a 61 c 61 g 172 t 511 others  
ORIGIN

Query Match 2.7%; Score 55.2; DB 17; Length 925;  
Best Local Similarity 13.6%; Pred. No. 0.052;  
Matches 51; Conservative 171; Mismatches 154; Indels 0; Gaps 0;

QY 1516 GCGCTTGTGTGGGAGACCGCGCGCGCTTGGCAAGCTCACTGAGCGCCTGAGAAGAGC 1575  
DB 924 SBSGSCSCSSBSSSSSMSTSSNSBSCSSSSBSSSSSTSSMSSSBSSSSSSSSSS 865  
QY 1576 GCGGAACGGCGCAACCTTTGGCGAGCGCGCGACGCCCTGAACGCCCTGAAGAGCTG 1635  
DB 864 SGTSSACVKNASSSCGCGCGMABCMCMSSSSSCGSASARGVKVRAISGAGKRGGS 805  
QY 1636 TACGAAGACCGCGCGCAGAGCTTACGACGACCGCCTTGGCAGGAGGACACTTCCGGAG 1695  
DB 804 GASASHSSSSACBSSSSSCASCMWASASSSSASRSRSGGAGGSGASSSSSSSSSA 745  
QY 1696 CAACAGGACCGCTGACGCTCCGGCAGCAAGGGCGGAGAGAGCGGCTTCCGCACTTGAA 1755  
DB 744 SAGSYVSSASSSSSSSVSCSSVASSMCSBSSSSASASSSSSSSASACASCCCT 685  
QY 1756 GCGCGCGAGCGCCCGAAGTGGCCCTTGACCAATGTTCCCGAAGACCGCGAGCTGAC 1815  
DB 684 SWSGSCSTASMSAARSSSSSSSSSSSMASASSSSSSSGSSSSSGSACBS 625  
QY 1816 CCGACCGGCCCTAAGTCGTGGTGGGGCGCGCTCAGTAGACGACAAAGCGCGTTCGTC 1875  
DB 624 MSSGGGGSVSASSGMSVSSGSGGSGGCGVGGSSSSSGSGSGSGSGSVSCSS 565  
QY 1876 GGCTCTCTGTAGACA 1891  
DB 564 GCMCRSCSSAAAAA 549

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 17:41:53 ; Search time 107 Seconds  
(without alignments)  
5821.126 Million cell updates/sec

5821.126 Million cell updates/sec

Title: US-09-940-550A-10

Perfect score: 2031

Sequence: 1 atggcacaagggtgtgtgac.....cgggaagacgtagcgcgtag 2031

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1796.6	88.5	2091	6 5190871-3	Patent No. 5190871
2	1793.4	88.3	3401	6 5190871-1	Patent No. 5190871
3	195	9.6	2633	1 US-08-452-267-2	Sequence 2, Appli
4	195	9.6	2633	3 US-09-123-644-2	Sequence 2, Appli
5	195	9.6	5534	1 US-08-452-267-3	Sequence 3, Appli
6	195	9.6	5534	3 US-09-123-644-3	Sequence 3, Appli
7	194	9.6	1701	4 US-09-080-625-1	Sequence 1, Appli
8	194	9.6	1701	4 US-09-080-625-1	Sequence 1, Appli
9	194	9.6	3336	4 US-09-080-625-2	Sequence 2, Appli
10	194	9.6	3336	4 US-09-080-625-2	Sequence 2, Appli
11	194	9.6	3877	4 US-09-080-625-4	Sequence 4, Appli
12	194	9.6	3877	4 US-09-080-625-4	Sequence 4, Appli
13	190.2	9.4	5560	3 US-08-817-188-5	Sequence 5, Appli
14	190	9.4	1034	4 US-09-367-293-2	Sequence 2, Appli
15	51.2	2.5	19124	2 US-08-487-8268-13	Sequence 13, Appli
16	51	2.5	1926	4 US-09-249-585A-4	Sequence 4, Appli
17	51	2.5	1931	2 US-09-130-114-2	Sequence 2, Appli
18	46.8	2.3	19124	2 US-08-487-8268-13	Sequence 13, Appli
19	46.6	2.3	11485	4 US-09-410-464-9	Sequence 9, Appli
20	45	2.2	722	4 US-08-998-416-780	Sequence 780, App
21	45	2.2	727	4 US-08-998-416-1011	Sequence 1011, Ap
22	44.8	2.2	8700	2 US-08-392-625-16	Sequence 16, Appli
23	44.8	2.2	8700	2 US-08-466-961A-16	Sequence 16, Appli
24	44.8	2.2	8700	2 US-08-645-193B-18	Sequence 18, Appli
25	44.4	2.2	5923	4 US-09-064-922-3	Sequence 3, Appli
26	44	2.2	2960	3 US-08-913-842-3	Sequence 3, Appli
27	43.4	2.1	1850	3 US-08-617-860B-32	Sequence 32, Appli

28	43.4	2.1	4098	2 US-08-605-106-4	Sequence 4, Appli
29	43	2.1	5852	1 US-07-867-106-2	Sequence 2, Appli
30	42.6	2.1	6265	4 US-09-129-112-3	Sequence 3, Appli
31	42.2	2.1	1431	4 US-09-316-083-2	Sequence 2, Appli
32	42	2.1	615	4 US-08-998-416-186	Sequence 186, App
33	41.8	2.1	7218	1 US-08-232-463-14	Sequence 14, Appli
34	41.4	2.0	7815	4 US-09-102-528-28	Sequence 28, Appli
35	41.4	2.0	20235	3 US-07-642-734C-3	Sequence 3, Appli
36	41.4	2.0	20235	3 US-08-439-009A-3	Sequence 3, Appli
37	41.4	2.0	168575	4 US-09-426-290-1	Sequence 1, Appli
38	41.4	2.0	4403765	4 US-09-103-840A-2	Sequence 2, Appli
39	41	2.0	2960	3 US-08-913-842-3	Sequence 3, Appli
40	40.8	2.0	3618	1 US-07-872-678A-36	Sequence 36, Appli
41	40.8	2.0	4411529	4 US-09-103-840A-1	Sequence 1, Appli
42	40.6	2.0	1150	1 US-08-161-406-1	Sequence 1, Appli
43	40.4	2.0	1600	4 US-09-434-288-10	Sequence 10, Appli
44	40.2	2.0	3196	2 US-09-096-982-4	Sequence 4, Appli
45	40.2	2.0	3196	2 US-08-653-650A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1	5190871-3	Patent No. 5190871
APPLICANT: COX, KAREN L.; KUHSTOSS, STUART A.; RAO, R. NAGARAJA		
RICHARDSON, MARK A.; SCHONER, BRIGITTE E.; SENO, EUGENE T.		
TITLE OF INVENTION: USE OF THE SITE-SPECIFIC INTEGRATING		
FUNCTION OF PHAGE C31		
NUMBER OF SEQUENCES: 3		
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: US/07/364, 959		
FILING DATE: 12-JUN-1989		
SEQ ID NO: 3:		
LENGTH: 2091		
5190871-3		
Query Match	88.5%	Score 1796.6; DB 6; Length 2091;
Best Local Similarity	99.28;	Pred. No. 0;
Matches 1805; Conservative	0;	Mismatches 14; Indels 0; Gaps 0;
QY 213 GATGTGACAGTACGGGGTGTGCTTACGACCGTACGCGGCGAGGAGATACAGTGC		272
DB 254 GGTGACACGTACGGGGTGTGCTTACGACCGTACGCGGCGAGGAGATTCGAGCGC		313
QY 273 AGCAAGCCGACGACACACAGCGTAGCGCCCAAGCAAGAGGCGCGACCTTACGCGCA		332
DB 314 AGCAAGCCGACGACACACAGCGTAGCGCCCAAGCAAGAGGCGCGACCTTACGCGCA		373
QY 333 AGTCGAGCGGAGCGGGGGCGGTTGAGGTTGCGGGCATTTGAGGAGCGCGCGCAC		392
DB 374 AGTCGAGCGGAGCGGGGGCGGTTGAGGTTGCGGGCATTTGAGGAGCGCGCGCAC		433
QY 393 GTCGCGTTGCGGAGCGGCGGCGGAGCGGCGGAGTTGAGCGCATCTGAACGATGCGCGC		452
DB 434 GTCGCGTTGCGGAGCGGCGGAGCGGCGGAGTTGAGCGCATCTGAACGATGCGCGC		493
QY 453 CGGCGGCTCAACATGATCATTTGTCTATGACGTGTGCGGCTTCTCGGCGCTGAAGTTCAT		512
DB 494 CGGCGGCTCAACATGATCATTTGTCTATGACGTGTGCGGCTTCTCGGCGCTGAAGTTCAT		553
QY 513 GGACGCGATTCGATTTGCTCTCGGAATTCCTGCGGCGGCGGCGGCGGCGGCGGCGG		572
DB 554 GGACGCGATTCGATTTGCTCTCGGAATTCCTGCGGCGGCGGCGGCGGCGGCGGCGG		613
QY 573 GGAAGGCGTCTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG		632
DB 614 GGAAGGCGTCTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG		673
QY 633 CGCGTGCACAAAGAAATCTTCTGCTGAAGTTCGCGGAGGAGGAGGAGGAGGAGGAGG		692
DB 674 CGCGTGCACAAAGAAATCTTCTGCTGAAGTTCGCGGAGGAGGAGGAGGAGGAGGAGG		733

QY 693 GCGCAATTGGGCGGTACGTGCGGGGAAGCGCCCTTACGGCTTCGAGCTTGTTCGGA 752  
 Db 734 GCGCAATTGGGCGGTACGTGCGGGGAAGCGCCCTTACGGCTTCGAGCTTGTTCGGA 793  
 QY 753 GACGAAGGAGATGACGCGCAACGCGCGAATGGTCAATGTGTCATCAACAAGTTAGCGCA 812  
 Db 794 GACGAAGGAGATGACGCGCAACGCGCGAATGGTCAATGTGTCATCAACAAGTTAGCGCA 853  
 QY 813 CTGACCACTCCCTTACCGGACCTTCGAGTTCGAGCCCGACGTAATCCGCTGTGTG 872  
 Db 854 CTGACCACTCCCTTACCGGACCTTCGAGTTCGAGCCCGACGTAATCCGCTGTGTG 913  
 QY 873 GCGTAGATCAAGACGACACAACACCTTCCCTTCAAGCCGGGACGTCAGCCGCAATCA 932  
 Db 914 GCGTAGATCAAGACGACACAACACCTTCCCTTCAAGCCGGGACGTCAGCCGCAATCA 973  
 QY 933 CCGGGGACATCAGCGGGCTTGTAGCGCATGAGCGCTGACGCGCTGCGGACCGGG 992  
 Db 974 CCGGGGACATCAGCGGGCTTGTAGCGCATGAGCGCTGACGCGCTGCGGACCGGG 1033  
 QY 993 CGAGACATTTGGGAAGAACCGCTTCAAGCGCTTGGAGCCGCGCAACCGTTATCGAAT 1052  
 Db 1034 CGAGACATTTGGGAAGAACCGCTTCAAGCGCTTGGAGCCGCGCAACCGTTATCGAAT 1093  
 QY 1053 CCTTCGGGACCGCGTATTCGGGCTTCGCGCTGAGGTATCTACAGAAGACCGGA 1112  
 Db 1094 CCTTCGGGACCGCGTATTCGGGCTTCGCGCTGAGGTATCTACAGAAGACCGGA 1153  
 QY 1113 CGGCAAGCGGACGACGAAGATTGAGGCTTACCGCATTCAGCGCGACCGCATCAGCTCCG 1172  
 Db 1154 CGGCAAGCGGACGACGAAGATTGAGGCTTACCGCATTCAGCGCGACCGCATCAGCTCCG 1213  
 QY 1173 GCGGTCGAGCTTGAATTCGAGACCATCATCGAGCCCGCTGAGTGTATGAGCTTCAGGC 1232  
 Db 1214 GCGGTCGAGCTTGAATTCGAGACCATCATCGAGCCCGCTGAGTGTATGAGCTTCAGGC 1273  
 QY 1233 GTGTTGAGCGGCAAGGGGCGCGCAAGGGGCTTCCCGGGCAAGCCATTGTTCGCG 1292  
 Db 1274 GTGTTGAGCGGCAAGGGGCGCGCAAGGGGCTTCCCGGGCAAGCCATTGTTCGCG 1333  
 QY 1293 CATGACAACTGTACTGCGAGTGTGGCGCGCTTCATGACTTCGAAGCGCGGGAAGATC 1352  
 Db 1334 CATGACAACTGTACTGCGAGTGTGGCGCGCTTCATGACTTCGAAGCGCGGGAAGATC 1393  
 QY 1353 GATCAAGGACTCTTACCGCTGCGCTCGCGGAAGGTGTCGACCGCTCCGACCTGGGCA 1412  
 Db 1394 GATCAAGGACTCTTACCGCTGCGCTCGCGGAAGGTGTCGACCGCTCCGACCTGGGCA 1453  
 QY 1413 GCAGAAAGGACGTGCAACGTGACATGGGGGCACTGACAAAGTTGTTGCGAAGCAT 1472  
 Db 1454 GCAGAAAGGACGTGCAACGTGACATGGGGGCACTGACAAAGTTGTTGCGAAGCAT 1513  
 QY 1473 CTTCAACAAGATCAGGCAAGCGGGAAGCGGCAAGAGACGTTGGCGCTTGTGGGAAGC 1532  
 Db 1514 CTTCAACAAGATCAGGCAAGCGGGAAGCGGCAAGAGACGTTGGCGCTTGTGGGAAGC 1573  
 QY 1533 CGCCCGACGCTTCGCGCAAGCTCACTGAGGGCGCTGAGAAAGAGCGGGAAGCGGCAACT 1592  
 Db 1574 CGCCCGACGCTTCGCGCAAGCTCACTGAGGGCGCTGAGAAAGAGCGGGAAGCGGCAACT 1633  
 QY 1593 TGTTCGGGAGCGCGCGGACGCTGAAAGCGCTTGAAGAGCTGTACGAAGACCGCGGCG 1652  
 Db 1634 TGTTCGGGAGCGCGCGGACGCTGAAAGCGCTTGAAGAGCTGTACGAAGACCGCGGCG 1693  
 QY 1653 AGGAGCTTACGAGGACCGCTTGGGCAAGAGCACTTCGGAAGCAAGGCGGCTGAC 1712  
 Db 1694 AGGAGCTTACGAGGACCGCTTGGGCAAGAGCACTTCGGAAGCAAGGCGGCTGAC 1753  
 QY 1713 GCTCCGCGACGAAGGGGCGGGAAGAGCGGCTTGGCAACTTGAAGCGCGGGAAGCGCGAA 1772  
 Db 1754 GCTCCGCGACGAAGGGGCGGGAAGAGCGGCTTGGCAACTTGAAGCGCGGGAAGCGCGAA 1813

QY 1773 GTTCCCTTGAACCAATGTTCCCGGAAGACGCGCCGACGCTGACCCGAGCGCCCTAAGTC 1832  
 Db 1814 GCTTCCCTTGAACCAATGTTCCCGGAAGACGCGCCGACGCTGACCCGAGCGCCCTAAGTC 1873  
 QY 1833 GTTGGGCGCGCGCTCAGTAGACGACAGCGCGTGTTCGTCGGGCTTCTGTAGACAA 1892  
 Db 1874 GTTGGGCGCGCGCTCAGTAGACGACAGCGCGTGTTCGTCGGGCTTCTGTAGACAA 1933  
 QY 1893 GATGTTGTCAGGAAGTGCATACGAGGCGGACGAGGCGGACGAGAGACGAGCGCGC 1952  
 Db 1934 GATGTTGTCAGGAAGTGCATACGAGGCGGACGAGGCGGACGAGAGACGAGCGCGC 1993  
 QY 1953 TTGATCAGTGGCGGAAGCGCGGACGAGCGGACGAGAGACGAGCGCGC 2012  
 Db 1994 TTGATCAGTGGCGGAAGCGCGGACGAGCGGACGAGAGACGAGCGCGC 2053  
 QY 2013 GGAAGCGTAGCGCGCTAG 2031  
 Db 2054 GGAAGCGTAGCGCGCTAG 2072

RESULT 2  
 5190871-1  
 ; Patent No. 5190871  
 ; APPLICANT: COX, KAREN L.; KUHSTOSS, STUART A.; RAO, R. NAGARAJA  
 ; RICHARDSON, MARK A.; SCHONER, BRIGITTE E.; SENDO, EUGENE T.  
 ; TITLE OF INVENTION: USE OF THE SITE-SPECIFIC INTEGRATING  
 ; FUNCTION OF PHAGE C31  
 ; NUMBER OF SEQUENCES: 3  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/364,959  
 ; FILING DATE: 12-JUN-1989  
 ; SEQ ID NO:1:  
 ; LENGTH: 3401  
 ; 5190871-1

Query Match. 88.3%; Score 1793.4; DB 6; Length 3401;  
 Best Local Similarity 99.1%; Pred. No. 0;  
 Matches 1803; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
 QY 213 GATGTGACAGGTACGCGGGTGTTCAGACCGTCAAGTTCGCGGACGCGGAGATAGCAGTGC 272  
 Db 532 GGTGACACAGTACGCGGGTGTTCAGACCGTCAAGTTCGCGGACGCGGAGATAGCAGTGC 272  
 QY 273 ACCAAGCCGACGACACAGCGGTAGCGCCCAAGAAAGAGAGGCGGCGGACCTTCAGCGCGA 332  
 Db 592 AGCAAGCCGACGACACAGCGGTAGCGCCCAAGAAAGAGAGGCGGCGGACCTTCAGCGCGA 332  
 QY 333 AGTCAGCGCGACGCGGCGGCTTCAAGTTCGTCGCGCATTTACGCGAAGCGCGGCGAC 392  
 Db 652 AGTCAGCGCGACGCGGCGGCTTCAAGTTCGTCGCGCATTTACGCGAAGCGCGGCGAC 392  
 QY 393 GTGCGGCTTCGCGGACGCGGCGGCGGCTTCAAGTTCGTCGCGCATTTACGCGAAGCGCGGCG 452  
 Db 712 GTGCGGCTTCGCGGACGCGGCGGCGGCTTCAAGTTCGTCGCGCATTTACGCGAAGCGCGGCG 452  
 QY 453 CGGCGGCTCAACATGATGATGCTATGAGCTGTGCGGCTTTCGCGGCTTTCGCGGCTTTCG 512  
 Db 772 CGGCGGCTCAACATGATGATGCTATGAGCTGTGCGGCTTTCGCGGCTTTCGCGGCTTTCG 512  
 QY 513 GAGCGGATTCGCGATGCTGCGAATGCTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 572  
 Db 832 GAGCGGATTCGCGATGCTGCGAATGCTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 572  
 QY 832 GAGCGGATTCGCGATGCTGCGAATGCTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 891  
 Db 891 GAGCGGATTCGCGATGCTGCGAATGCTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCG 891  
 QY 892 GGAAGCGGCTTCGCGGAGGAAAGCTATGAGCTGATTCACCTGATTCAGCTGATTCAG 951  
 Db 951 GGAAGCGGCTTCGCGGAGGAAAGCTATGAGCTGATTCACCTGATTCAGCTGATTCAG 951  
 QY 951 GCGGTCGACAAAGATCTTCGCTGAAGTTCGCGGAGATTCGACACGAGAAACCTTCA 992  
 Db 992 GCGGTCGACAAAGATCTTCGCTGAAGTTCGCGGAGATTCGACACGAGAAACCTTCA 992  
 QY 992 GCGGTCGACAAAGATCTTCGCTGAAGTTCGCGGAGATTCGACACGAGAAACCTTCA 1011  
 Db 1011 GCGGTCGACAAAGATCTTCGCTGAAGTTCGCGGAGATTCGACACGAGAAACCTTCA 1011  
 QY 693 GCGGCAATTGGGCGGTACGTGCGGGGAAGGCGCCTTACGGCTTCGAGCTTGTTCGGA 752





QY 90 ATATAATATTTCAAATATTTTTCACAAATAAAGAATGTAGTATATAGCAATTGCTTTT 149  
|||||  
Db 753 ATATAATATTTCAAATATTTTTCACAAATAAAGAATGTAGTATATAGCAATTGCTTTT 812  
QY 150 CTGTAGTTTAAAGTGTATATTTTAACTTTCTAATATATGACCAAAATTTT 209  
|||||  
Db 813 CTGTAGTTTAAAGTGTATATTTTAACTTTCTAATATATGACCAAAATTTT 872  
QY 210 GTTGATGTGCAGTA 224  
|||||  
Db 873 GTTGATGTGCAGTA 887

## RESULT 4

US-09-123-644-2  
; Sequence 2, Application US/09123644  
; Patent No. 6127606  
; GENERAL INFORMATION:  
; APPLICANT: Bennett, Malcolm  
; APPLICANT: May, Sean  
; APPLICANT: Ramsay, Nicola  
; TITLE OF INVENTION: Method of Using Transactivation Proteins to  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Barnes & Thornburg  
; STREET: 11 South Meridian  
; CITY: Indianapolis  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/123,644  
; FILING DATE: 28-JUL-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Breen, John P.  
; REGISTRATION NUMBER: 38,833  
; REFERENCE/DOCKET NUMBER: 6653-60788  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317) 231-7745  
; TELEFAX: (317) 231-7433  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2633 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: pUMIGIT(insert)  
; US-09-123-644-2

Query Match 9.6%; Score 195; DB 3; Length 2633;  
Best Local Similarity 100.0%; Pred. No. 2.8e-38;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 TACGTAAGTTTCTGCTTCACTTTGATATATATATAATTAATCAATTAAGTAGTA 89  
|||||  
Db 693 TACGTAAGTTTCTGCTTCACTTTGATATATATATAATTAATCAATTAAGTAGTA 752  
QY 90 ATATAATATTTCAAATATTTTTCACAAATAAAGAATGTAGTATATGCAATGCTTTT 149  
|||||  
Db 753 ATATAATATTTCAAATATTTTTCACAAATAAAGAATGTAGTATATGCAATGCTTTT 812  
QY 150 CTGTAGTTTAAAGTGTATATTTTAACTTTCTAATATATGACCAAAATTTT 209  
|||||

Db 813 CTGTAGTTTAAAGTGTATATTTTAACTTTCTAATATATGACCAAAATTTT 872  
QY 210 GTTGATGTGCAGTA 224  
|||||  
Db 873 GTTGATGTGCAGTA 887

## RESULT 5

US-08-452-267-3  
; Sequence 3, Application US/08452267  
; Patent No. 5801027  
; GENERAL INFORMATION:  
; APPLICANT: Bennett, Malcolm  
; APPLICANT: May, Sean  
; APPLICANT: Ramsay, Nicola  
; TITLE OF INVENTION: Control of Genes in Transgenic Plants  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Barnes & Thornburg  
; STREET: 11 South Meridian  
; CITY: Indianapolis  
; STATE: IN  
; COUNTRY: USA  
; ZIP: 46204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/452,267  
; FILING DATE: 26-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Breen, John P.  
; REGISTRATION NUMBER: 38,833  
; REFERENCE/DOCKET NUMBER: 6653-25744  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317) 231-7745  
; TELEFAX: (317) 231-7433  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5534 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; IMMEDIATE SOURCE:  
; CLONE: pUMIGIT  
; US-08-452-267-3

Query Match 9.6%; Score 195; DB 1; Length 5534;  
Best Local Similarity 100.0%; Pred. No. 3.9e-38;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 TACGTAAGTTTCTGCTTCACTTTGATATATATATAATTAATCAATTAAGTAGTA 89  
|||||  
Db 688 TACGTAAGTTTCTGCTTCACTTTGATATATATATAATTAATCAATTAAGTAGTA 747  
QY 90 ATATAATATTTCAAATATTTTTCACAAATAAAGAATGTAGTATATGCAATGCTTTT 149  
|||||  
Db 748 ATATAATATTTCAAATATTTTTCACAAATAAAGAATGTAGTATATGCAATGCTTTT 807  
QY 150 CTGTAGTTTAAAGTGTATATTTTAACTTTCTAATATATGACCAAAATTTT 209  
|||||  
Db 808 CTGTAGTTTAAAGTGTATATTTTAACTTTCTAATATATGACCAAAATTTT 867  
QY 210 GTTGATGTGCAGTA 224  
|||||  
Db 868 GTTGATGTGCAGTA 882

RESULT 6

US-09-123-644-3

; Sequence 3, Application US/09123644

; Patent No. 6127606

; GENERAL INFORMATION:

; APPLICANT: Bennett, Malcolm

; APPLICANT: May, Sean

; APPLICANT: Ramsay, Nicola

; TITLE OF INVENTION: Method of Using Transactivation Proteins to

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Barnes & Thornburg

; STREET: 11 South Meridian

; CITY: Indianapolis

; STATE: IN

; COUNTRY: USA

; ZIP: 46204

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/123,644

; FILING DATE: 28-JUL-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Breen, John P.

; REGISTRATION NUMBER: 38,833

; REFERENCE/DOCKET NUMBER: 6653-60788

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (317) 231-7745

; TELEFAX: (317) 231-7433

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5534 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: circular

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; IMMEDIATE SOURCE:

; CLONE: pUMIGIT

; US-09-123-644-3

Query Match 9.6%; Score 195; DB 3; Length 5534;  
Best Local Similarity 100.0%; Pred. NO. 3.9e-38;  
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 30 TACGTAAGTTCTGCTTCTACCTTGATATATATAATAATTATCATTAATAGTAGTA 89  
Db 688 TACGTAAGTTCTGCTTCTACCTTGATATATATAATAATTATCATTAATAGTAGTA 747  
OY 90 ATATAATATTTCAATATTTTTCAAAATAAAGATGTAGTATAGCAATTGCTTTT 149  
Db 748 ATATAATATTTCAATATTTTTCAAAATAAAGATGTAGTATAGCAATTGCTTTT 807  
OY 150 CTGTAAGTTTATAGTGTGTATATTTTAATTATTAATCTTTCTAATATATAGCAAAATTT 209  
Db 808 CTGTAAGTTTATAGTGTGTATATTTTAATTATTAATCTTTCTAATATATAGCAAAATTT 867  
OY 210 GTTGATGTGACAGTA 224  
Db 868 GTTGATGTGACAGTA 882

RESULT 7

US-09-080-625-1

; Sequence 1, Application US/09080625

; Patent No. 6307123

; GENERAL INFORMATION:

APPLICANT: Kriz, Alan L.

APPLICANT: Spencer, T. Michael

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE

TITLE OF INVENTION: IDENTIFICATION

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: TX

COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/080,625

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Hanson, Robert E.

REGISTRATION NUMBER: P-42,628

REFERENCE/DOCKET NUMBER: DEKM:161

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 418-3000

TELEFAX: (512) 474-7577

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1701 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-080-625-1

Query Match 9.6%; Score 194; DB 4; Length 1701;  
Best Local Similarity 97.5%; Pred. NO. 4e-38;  
Matches 197; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 23 GGTGATACGTAAGTTCTGCTTCTACCTTTGATATATATAATAATTATCATTAATT 82  
Db 370 GAGCTGAAGTAAGTTCTGCTTCTACCTTTGATATATATAATAATTATCATTAATT 429  
OY 83 AGTAGAATATATATTTTCAAAATATTTTTCAAAATAAAGATGTAGTATATAGCAAT 142  
Db 430 AGTAGAATATATATTTTCAAAATATTTTTCAAAATAAAGATGTAGTATATAGCAAT 489  
OY 143 TGCTTTTCTGTAAGTTTATAGTGTATATTTTAATTATTAATCTTTCTAATATATGACC 202  
Db 490 TGCTTTTCTGTAAGTTTATAGTGTATATTTTAATTATTAATCTTTCTAATATATGACC 549  
OY 203 AAAATTTGTGATGTGACAGTA 224  
Db 550 AAAATTTGTGATGTGACAGTA 571

RESULT 8

US-09-695-782-1

; Sequence 1, Application US/09695782

; Patent No. 6433252

; GENERAL INFORMATION:

; APPLICANT: KRIZ, ALAN L.

; APPLICANT: GRIFFOR, MATTHEW

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE IDENTIFICATION

FILE REFERENCE: DEKM:161USCI

CURRENT APPLICATION NUMBER: US/09/695,782

PRIOR FILING DATE: 2000-10-24

PRIOR APPLICATION NUMBER: 09/080,625

PRIOR FILING DATE: 1998-05-18

NUMBER OF SEQ ID NOS: 19

SOFTWARE: Patentln Ver. 2.0

; SEQ ID NO 1



US-09-080-625-4

1 CORRESPONDENCE ADDRESS:  
2 ADDRESSEE: Arnold White & Durkee  
3 STREET: P.O. Box 4433  
4 CITY: Houston  
5 STATE: TX  
6 COUNTRY: USA  
7 ZIP: 77210-4433  
8  
9 COMPUTER READABLE FORM:  
10 MEDIUM TYPE: Floppy disk  
11 COMPUTER: IBM PC compatible  
12 OPERATING SYSTEM: PC-DOS/MS-DOS  
13 SOFTWARE: PatentIn Release #1.0, Version #1.30  
14  
15 CURRENT APPLICATION DATA:  
16 APPLICATION NUMBER: US/09/080,625  
17  
18 FILING DATE:  
19  
20 CLASSIFICATION:  
21 ATTORNEY/AGENT INFORMATION:  
22 NAME: Hanson, Robert E.  
23 REGISTRATION NUMBER: P-42,628  
24 REFERENCE/DOCKET NUMBER: DEKM:161  
25 TELECOMMUNICATION INFORMATION:  
26 TELEPHONE: (512) 418-3000  
27 TELEFAX: (512) 474-7577  
28 INFORMATION FOR SEQ ID NO: 4:  
29 SEQUENCE CHARACTERISTICS:  
30 LENGTH: 3877 base pairs  
31 TYPE: nucleic acid  
32 STRANDEDNESS: double  
33 TOPOLOGY: linear

```

Query Match          9.68; Score 194; DB 4; Length 3877;
Best Local Similarity 97.58; Pred. No. 5.8e-38;
Matches 197; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 23 GGGTGGATACGTAAGTTTCGCTCTACCTTGATATATATATAATTATCATTAATT 82
   || || | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1982 GAGCTGAAGGTAAGTTTCGCTCTACCTTGATATATATATAATTATCATTAATT 2041

QY 83 AGTAGTAATATAATATTTCAAAATATTTTTCAAAATAAAGAACTAGTATATAGCAAT 142
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2042 AGTAGTAATATAATATTTCAAAATATTTTTCAAAATAAAGAACTAGTATATAGCAAT 2101

QY 143 TGCTTTTCTGTAGTTTATTAAGTGTGTATATTTTAATTATACTTTTCTAATATATGACC 202
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2102 TGCTTTTCTGTAGTTTATTAAGTGTGTATATTTTAATTATACTTTTCTAATATATGACC 2161

QY 203 AAAATTGTTGATGTCAGCTA 224
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2162 AAAATTGTTGATGTCAGCTA 2183

RESULT 12
US-09-695-782-4
; Sequence 4, Application US/09695782
; Patent No. 6433252
; GENERAL INFORMATION:
; APPLICANT: KRIZ, ALAN L.
; APPLICANT: GRIFFOR, MATTHEW
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TRANSGENE IDENTIFICATION
; FILE REFERENCE: DEKM:161USCI
; CURRENT APPLICATION NUMBER: US/09/695,782
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: 09/080,625
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 3877
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic

```

[illegible]

```

RESULT 13
US-08-817-188-5/c
; Sequence 5, Application US/08817188
; Patent No. 6074876
; GENERAL INFORMATION:
; APPLICANT: DE BLOCK, MARC
; TITLE OF INVENTION: GENETIC TRANSFORMATION USING A PARP INHIBITOR
; FILE REFERENCE: 2121-0127P
; CURRENT APPLICATION NUMBER: US/08/817,188
; CURRENT FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: PCT/EP96/03366
; EARLIER FILING DATE: 1996-07-31
; EARLIER APPLICATION NUMBER: EP 95401844.6
; EARLIER FILING DATE: 1995-08-04
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 5560
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: T-DNA of
; OTHER INFORMATION: plasmid pTHW142
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(25)
; OTHER INFORMATION: RB: right border sequence of octopine TL-DNA from
; OTHER INFORMATION: pTiB6S3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84)..(296)
; OTHER INFORMATION: 3' g7: 3' untranslated region containing the
; OTHER INFORMATION: polyadenylation signal of gene 7 of Agrobacterium
; OTHER INFORMATION: T-DNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (318)..(869)
; OTHER INFORMATION: bar: region coding for phosphinotricin
; OTHER INFORMATION: acetyltransferase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (830)..(2760)
; OTHER INFORMATION: PSSU: promoter region of Rubisco small subunit
; OTHER INFORMATION: gene of Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2765)..(3058)
; OTHER INFORMATION: 3' untranslated region of the CamV 35S transcript
; OTHER INFORMATION: containing polyadenylation signals

```



Oliver...

Best Local Similarity: 9.48; Score: 100

[illegible]

RESULT 14  
 US-09-367-293-2  
 Sequence 2, Application US/09367293  
 Patent No. 6444878  
 GENERAL INFORMATION:  
 APPLICANT: Donaldson, Lala A.  
 APPLICANT: Bojsen, Kirsten  
 APPLICANT: Jorgensen, Kirsten  
 TITLE OF INVENTION: Jorsboe, Morten  
 FILE REFERENCE: SELECTION METHOD FOR TRANSGENIC PLANTS  
 CURRENT APPLICATION NUMBER: DYO021.001APC  
 PRIOR FILING DATE: 1999-12-23  
 PRIOR FILING DATE: 1999-12-23  
 PRIOR FILING DATE: 1998-02-05  
 PRIOR APPLICATION NUMBER: PCT/GB98/00367  
 PRIOR FILING DATE: 1997-02-07  
 GB 9702592.8

OTHER INFORMATION:  
US-09-367-293-2

Query Match	Best Local Similarity	Matches	Score	Protein
9.48	100	100	100	Phosphate deaminase

	33	GTAGTTTCTGCTTCTACCTTGCATATATATATAATAATATCATTAATTAGTAGTAATA	92
QY	33	GTAGTTTCTGCTTCTACCTTGCATATATATATAATAATATCATTAATTAGTAGTAATA	92
Db	235	GTAGTTTCTGCTTCTACCTTGCATATATATATAATAATATCATTAATTAGTAGTAATA	92
QY	93	TAATATTTCAAAATATATTTTTTCAAAATAAAAGAATGTAGTATAGCAATTGCTTTCTG	152
Db	295	TAATATTTCAAAATATATTTTTTCAAAATAAAAGAATGTAGTATAGCAATTGCTTTCTG	152
QY	153	TAGTTTATAAGTGCTGTATATTTTTTAATTTATAACTTTCTTAATATATGACCAAAATTTGTT	212
Db	355	TAGTTTATAAGTGCTGTATATTTTTTAATTTATAACTTTCTTAATATATGACCAAAATTTGTT	212
QY	213	GATGTGCAGG 222	
Db	415	GATGTGCAGG 424	

RESULT 15  
US-08-487

RESULT 15  
US-08-487-826D

Sequence 13, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION  
Appr

APPLICANT: SIM, Kim L.  
 APPLICANT: CHITNIS, Chetan  
 APPLICANT: MILLER, Louis H.  
 APPLICANT: PETERSON, David S.  
 APPLICANT: SU, Xin-zhan  
 TITLE OF INVENTION: Thomas E.  
 NUMBER OF INVENTION: BINDING DOMAINS  
 CORRESPONDENCE ADDRESSES: AND PLASMODIUM  
 ADDRESSEE: 45 FALCIPARUM  
 STREET: KNOX, ERYTHROCYTE

... BINDING PROTEINS

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
CURRENT APPLICATION: Patent In Release #1.0, Version #1.25  
APPLICATION NUMBER:  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NI121.001CPI  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS  
LENGTH: 19124 base pairs







Db 181 ATAACCTTTCTAATATATATGACCAAAATTTGTTGATGTGACAGGTACGGGGTGTCTACGAC 240  
QY 241 CGTCAGTCGGCGGAGCGCGAGAAATAGCAGTGCAGCAAGCCAGCGACACAGCGTAGCGCC 300  
Db 241 CGTCAGTCGGCGGAGCGCGAGAAATAGCAGTGCAGCAAGCCAGCGACACAGCGTAGCGCC 300  
QY 301 AACGAAGACAAAGCGCGCGACCTTCAGCGCGAGAGTCGAGCGGAGCGGCGCGGTTCAAG 360  
Db 301 AACGAAGACAAAGCGCGCGACCTTCAGCGCGAGAGTCGAGCGGAGCGGCGCGGTTCAAG 360  
QY 361 TTCGTCGGGCAATTCAGCGCAAGCGCGGCGACGTGCGGCTTCGGGACGGCGGAGCGCCCG 420  
Db 361 TTCGTCGGGCAATTCAGCGCAAGCGCGGCGACGTGCGGCTTCGGGACGGCGGAGCGCCCG 420  
QY 421 GAGTTCGAACGCACTCTGAACGAATGCGCGCGCGCGGCGCTCAACATGATCATTTGTCTAT 480  
Db 421 GAGTTCGAACGCACTCTGAACGAATGCGCGCGCGCGGCGCTCAACATGATCATTTGTCTAT 480  
QY 481 GACGTGTGCGGCTTCTCGCGCTGAAAGTCATGAGCGCGATTCGATTTGTCTCGGAATTG 540  
Db 481 GACGTGTGCGGCTTCTCGCGCTGAAAGTCATGAGCGCGATTCGATTTGTCTCGGAATTG 540  
QY 541 CTCGCCCTGGGCGTGACGATGTGTTTCCACTCAGGAAGCGCTTCCGGCAGGGAAACGTC 600  
Db 541 CTCGCCCTGGGCGTGACGATGTGTTTCCACTCAGGAAGCGCTTCCGGCAGGGAAACGTC 600  
QY 601 ATGACCTGATTCACCTGATTTATGCGCGCTCGACGCGTCCGACAAAGATCTTCGCTGAAG 660  
Db 601 ATGACCTGATTCACCTGATTTATGCGCGCTCGACGCGTCCGACAAAGATCTTCGCTGAAG 660  
QY 661 TCGCGGAAGATCTCGACACGAAGAACTTCAGCGCGAAATGGCGGGTACGTCGGCGGG 720  
Db 661 TCGCGGAAGATCTCGACACGAAGAACTTCAGCGCGAAATGGCGGGTACGTCGGCGGG 720  
QY 721 AAGCGCGCTTACGGCTTCGAGCTGTTTTCGAGACGAAGAGATCAGCGCGCAACGCGCA 780  
Db 721 AAGCGCGCTTACGGCTTCGAGCTGTTTTCGAGACGAAGAGATCAGCGCGCAACGCGCA 780  
QY 781 ATGCTCAATGTCTCATCAACAAGTTAGCGCACTCGACCACTCCCTTACCGGACCTTC 840  
Db 781 ATGCTCAATGTCTCATCAACAAGTTAGCGCACTCGACCACTCCCTTACCGGACCTTC 840  
QY 841 GAGTTCGAGCGCGGACGTAATCGGGTGTGTTGCGGTGAGATCAAGACGCAACACCTT 900  
Db 841 GAGTTCGAGCGCGGACGTAATCGGGTGTGTTGCGGTGAGATCAAGACGCAACACCTT 900  
QY 901 CCCTTCAAGCCGGGAGTCAAGCCGCCATTACCCGGGCGAGCATCAGCGGGCTTTGTAAG 960  
Db 901 CCCTTCAAGCCGGGAGTCAAGCCGCCATTACCCGGGCGAGCATCAGCGGGCTTTGTAAG 960  
QY 961 CGCATGAGCGCTGACGCGCGTCCGACCGCGGCGAGACGATTTGGAGAAGAACCGCTTCA 1020  
Db 961 CGCATGAGCGCTGACGCGCGTCCGACCGCGGCGAGACGATTTGGAGAAGAACCGCTTCA 1020  
QY 1021 AGCGCTGGGACCGCGCAACCTTATGCAATCCTTCGGGACCGCGTATTGCGGCGTTC 1080  
Db 1021 AGCGCTGGGACCGCGCAACCTTATGCAATCCTTCGGGACCGCGTATTGCGGCGTTC 1080  
QY 1081 GCCGCTGAGGTGATCTACAAGAAGAAGCGCGGAGCGGACGCGGACGGAAGATTGAGGT 1140  
Db 1081 GCCGCTGAGGTGATCTACAAGAAGAAGCGCGGAGCGGACGCGGACGGAAGATTGAGGT 1140  
QY 1141 TACCGCATTCAGCGGACCGCATCACGCTCCGGCGGTGAGCTTGATTGCGGACGATC 1200  
Db 1141 TACCGCATTCAGCGGACCGCATCACGCTCCGGCGGTGAGCTTGATTGCGGACGATC 1200  
QY 1201 ATCGAGCCCGCTGAGTGTATGAGCTTCAGGCGTGTGAGCGGACGAGGGCGCGCAAG 1260  
Db 1201 ATCGAGCCCGCTGAGTGTATGAGCTTCAGGCGTGTGAGCGGACGAGGGCGCGCAAG 1260  
QY 1261 GGGCTTCCCGGGGCAAGCATTCCTGCGCCATGAGCAAGCTGTACTGCGAGTGTGC 1320  
Db 1261 GGGCTTCCCGGGGCAAGCATTCCTGCGCCATGAGCAAGCTGTACTGCGAGTGTGC 1320

QY 1321 GCCGTATGACTTCGAAGCGCGGGGAGAATGATCAAGGACTCTTACCGCTGCGTGC 1380  
Db 1321 GCCGTATGACTTCGAAGCGCGGGGAGAATGATCAAGGACTCTTACCGCTGCGTGC 1380  
QY 1381 CGGAAGTGTGACCGCTCCGACCTGGGACGACGAAGGACAGTGCACGTCAGCATG 1440  
Db 1381 CGGAAGTGTGACCGCTCCGACCTGGGACGACGAAGGACAGTGCACGTCAGCATG 1440  
QY 1441 GCGGCACTCGACAAGTTCGTTGCGGGAACGCAATCTTCAACAAGATCAGGACCGCAAGGC 1500  
Db 1441 GCGGCACTCGACAAGTTCGTTGCGGGAACGCAATCTTCAACAAGATCAGGACCGCAAGGC 1500  
QY 1501 GACGAAGAGACGTTGGCGCTTCTGTGGGAAGCGCGCGGACGCTTCGCAAGCTCACTAG 1560  
Db 1501 GACGAAGAGACGTTGGCGCTTCTGTGGGAAGCGCGCGGACGCTTCGCAAGCTCACTAG 1560  
QY 1561 GCGGCTGAGAAGAGCGGCGGAACCGGCGGAACCTTGTGCGGAGCGCGCGACGCCCTGAAC 1620  
Db 1561 GCGGCTGAGAAGAGCGGCGGAACCGGCGGAACCTTGTGCGGAGCGCGCGACGCCCTGAAC 1620  
QY 1621 GCCCTGAAGAGCTGTACGAAGACCGCGCGGAGGAGCTTACGACGACCGCTTGCCAGG 1680  
Db 1621 GCCCTGAAGAGCTGTACGAAGACCGCGCGGAGGAGCTTACGACGACCGCTTGCCAGG 1680  
QY 1681 AAGCACTTCCGGAAGCAAGCAGGACGCGCTGACGCTCCGCGACGAAGGGCGGGAAGAGCGG 1740  
Db 1681 AAGCACTTCCGGAAGCAAGCAGGACGCGCTGACGCTCCGCGACGAAGGGCGGGAAGAGCGG 1740  
QY 1741 CTTGCCGAACCTTGAAGCGCGCGGAGCGCGGAGCGCGGAGAGTGGCCCTTGACCAATGTTCCCGGAA 1800  
Db 1741 CTTGCCGAACCTTGAAGCGCGCGGAGCGCGGAGCGCGGAGAGTGGCCCTTGACCAATGTTCCCGGAA 1800  
QY 1801 GACGCGGAGCTGACCGCGGACCGCGGCTTGTGTTGAGCAAGATCGTTGTCAGGAAGTGCAGTACGAGC 1860  
Db 1801 GACGCGGAGCTGACCGCGGACCGCGGCTTGTGTTGAGCAAGATCGTTGTCAGGAAGTGCAGTACGAGC 1860  
QY 1861 AAGCGGCTGTTCTGTCGGGCTCTTCGTAGACAAGATCGTTGTCAGGAAGTGCAGTACGAGC 1920  
Db 1861 AAGCGGCTGTTCTGTCGGGCTCTTCGTAGACAAGATCGTTGTCAGGAAGTGCAGTACGAGC 1920  
QY 1921 AGGGGCGAGGAAGCGCCCATCGAGAAGCGCGCTTCGATCAGCTGGGCGAAGCGCGGAC 1980  
Db 1921 AGGGGCGAGGAAGCGCCCATCGAGAAGCGCGCTTCGATCAGCTGGGCGAAGCGCGGAC 1980  
QY 1981 GACGACGACGAAGACGAGCGCCGAGGACGCGGACGGAAGAGCTAGCGGCTAG 2031  
Db 1981 GACGACGACGAAGACGAGCGCCGAGGACGCGGACGGAAGAGCTAGCGGCTAG 2031

RESULT 2  
US-09-940-550A-9  
; Sequence 9, Application US/09940550A  
; Publication No. US20030033635A1  
; GENERAL INFORMATION:  
; APPLICANT: MANKIN, LUKE  
; APPLICANT: MCKERSIE, BRYAN  
; TITLE OF INVENTION: SELF-EXCISING POLYNUCLEOTIDES AND USES THEREOF  
; FILE REFERENCE: 16313-0055  
; CURRENT APPLICATION NUMBER: US/09/940,550A  
; PRIOR FILING DATE: 2001-08-27  
; PRIOR APPLICATION NUMBER: 60/227,961  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 2031  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: C3Int DNA  
; OTHER INFORMATION: sequence  
US-09-940-550A-9



Query Match 99.2%; Score 2015; DB 9; Length 2031;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 2021; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ATGCACACAGGGGTGTGACCGGGGTGATACGTAGTTCTGCTTACCTTTGATATA 60  
Db 1 ATGCACACAGGGGTGTGACCGGGGTGATACGTAGTTCTGCTTACCTTTGATATA 60  
QY 61 TATATATATATATATATAGCAATGCTTTCTAGTTTAAATATTTTCAAAATA 120  
Db 61 TATATATATATATATATAGCAATGCTTTCTAGTTTAAATATTTTCAAAATA 120  
QY 121 AAAGATAGTATATAGCAATGCTTTCTAGTTTAAATATTTTCAAAATA 180  
Db 121 AAAGATAGTATATAGCAATGCTTTCTAGTTTAAATATTTTCAAAATA 180  
QY 181 ATAACTTTTCTAATATATGACCAAAATTTGTGATGTCAGGTACGGGGTGTACGAC 240  
Db 181 ATAACTTTTCTAATATATGACCAAAATTTGTGATGTCAGGTACGGGGTGTACGAC 240  
QY 241 CGTACGTCCGGCGAGCGCGCAATAGCAGTGCAGCAAGCCCGACACAGCGTAGCGCC 300  
Db 241 CGTACGTCCGGCGAGCGCGCAATAGCAGTGCAGCAAGCCCGACACAGCGTAGCGCC 300  
QY 301 AACGAGACACAGGGCGCGCACTTCAGCGCAAGTCGAGCGCGGGGGCGGCTTCAGG 360  
Db 301 AACGAGACACAGGGCGCGCACTTCAGCGCAAGTCGAGCGCGGGGGCGGCTTCAGG 360  
QY 361 TTGCTCGGGCATTTTCAGCGAGCGCGCGCACGTCGCGCTTCGGAGCGCGGAGCGCCG 420  
Db 361 TTGCTCGGGCATTTTCAGCGAGCGCGCGCACGTCGCGCTTCGGAGCGCGGAGCGCCG 420  
QY 421 GAGTTGGAACGCATCTGGAAGAAATGCGCGCGCGCGCGCTCAACATGATCATTTGCTAT 480  
Db 421 GAGTTGGAACGCATCTGGAAGAAATGCGCGCGCGCGCGCTCAACATGATCATTTGCTAT 480  
QY 481 GAGTGTGCGGCTTCTCGCGCTGAAGTTCATGACGCGGATTCGGATTTGTCGGAATTG 540  
Db 481 GAGTGTGCGGCTTCTCGCGCTGAAGTTCATGACGCGGATTCGGATTTGTCGGAATTG 540  
QY 541 CTCGCCCTGGGCGTGACGATTTTTCCTACTCAGAGAGCGCTTCCGGCAGGGAACGTC 600  
Db 541 CTCGCCCTGGGCGTGACGATTTTTCCTACTCAGAGAGCGCTTCCGGCAGGGAACGTC 600  
QY 601 ATGACCTGATTCACCTGATTAATGCGGCTCGACGCGTGCACAAAGAAATCTTCGCTGAAG 660  
Db 601 ATGACCTGATTCACCTGATTAATGCGGCTCGACGCGTGCACAAAGAAATCTTCGCTGAAG 660  
QY 661 TCGCGGAAGATTTCTGACACGAAGAACTTCAGCGCGAAATGGCGGTACGTGCGCGG 720  
Db 661 TCGCGGAAGATTTCTGACACGAAGAACTTCAGCGCGAAATGGCGGTACGTGCGCGG 720  
QY 721 AAGCGCCTTACGCTTCGAGCTTGTTCGAGACGAAGAGATCACGCGCAACGGCCGA 780  
Db 721 AAGCGCCTTACGCTTCGAGCTTGTTCGAGACGAAGAGATCACGCGCAACGGCCGA 780  
QY 781 ATGTCATATGTCGTCATCAACAAGTTAGCGCAGTCCACTCCCTTACCGGACCTTC 840  
Db 781 ATGTCATATGTCGTCATCAACAAGTTAGCGCAGTCCACTCCCTTACCGGACCTTC 840  
QY 841 GAGTTGAGCGCGAGCTTAATCCGGTGGTGGCTGAGATCAAGAGCGCAAAACACCTT 900  
Db 841 GAGTTGAGCGCGAGCTTAATCCGGTGGTGGCTGAGATCAAGAGCGCAAAACACCTT 900  
QY 901 CCCTTCAAGCGGGGAGTCAAGCGCCCATTCACCGGGGAGCATCACGGGCTTTGTAG 960  
Db 901 CCCTTCAAGCGGGGAGTCAAGCGCCCATTCACCGGGGAGCATCACGGGCTTTGTAG 960  
QY 961 CGCATGAGCGCTGACGCGCTGCGAGCGGGGAGAGCATTTGGGAAGACCGCTTCA 1020  
Db 961 CGCATGAGCGCTGACGCGCTGCGAGCGGGGAGAGCATTTGGGAAGACCGCTTCA 1020

QY 1021 AGCGCTGGGACCCGGCAACCGTTATCGGAATCCTTCGGGACCCGGTATTGCGGCTTC 1080  
Db 1021 AGCGCTGGGACCCGGCAACCGTTATCGGAATCCTTCGGGACCCGGTATTGCGGCTTC 1080  
QY 1081 GCCGTGAGTGTATCTACAAGAAGAGCCGAGCGCACCGGACCAAGAGATTGAGGGT 1140  
Db 1081 GCCGTGAGTGTATCTACAAGAAGAGCCGAGCGCACCGGACCAAGAGATTGAGGGT 1140  
QY 1141 TACCGCATTCAGCGCGGACCCGATCAGCTCCGGCGGTGAGCTTGATTGCGGACCGATC 1200  
Db 1141 TACCGCATTCAGCGCGGACCCGATCAGCTCCGGCGGTGAGCTTGATTGCGGACCGATC 1200  
QY 1201 ATGAGCCCGCTGAGTGTATGAGCTTCAGGCGGTGTTGACGGGCGCGGCAAG 1260  
Db 1201 ATGAGCCCGCTGAGTGTATGAGCTTCAGGCGGTGTTGACGGGCGCGGCAAG 1260  
QY 1261 GGGCTTCCGGGGGCAAGCCATTCGTCCGCCATGAGACAGCTGTACTGCGAGTGTGGC 1320  
Db 1261 GGGCTTCCGGGGGCAAGCCATTCGTCCGCCATGAGACAGCTGTACTGCGAGTGTGGC 1320  
QY 1321 GCCGTGACTTTCGAAGCGGGGGAAGATCGATCAAGAGCTTTACCGCTGCGCTGCG 1380  
Db 1321 GCCGTGACTTTCGAAGCGGGGGAAGATCGATCAAGAGCTTTACCGCTGCGCTGCG 1380  
QY 1381 CGGAAGTGTGACACCCGCTCCGACCTGGGCGAGCAGCAGAGCAGCTGCAACGTACAGATG 1440  
Db 1381 CGGAAGTGTGACACCCGCTCCGACCTGGGCGAGCAGCAGAGCAGCTGCAACGTACAGATG 1440  
QY 1441 GCGGCACTGACACAAGTTGCTGTCGGAAGCGCATCTTCAACAAGATCAGGACGCGCGAAGGC 1500  
Db 1441 GCGGCACTGACACAAGTTGCTGTCGGAAGCGCATCTTCAACAAGATCAGGACGCGCGAAGGC 1500  
QY 1501 GACGAAGAGAGCTTGCGCTTCTGTGGGAAGCGCGCGCGCTTCGGCAAGCTCAGTGAAG 1560  
Db 1501 GACGAAGAGAGCTTGCGCTTCTGTGGGAAGCGCGCGCGCTTCGGCAAGCTCAGTGAAG 1560  
QY 1561 GCGCTGAGAGAGCGGGCAACGGCGCAACCTTGTGCGGAGCGCGCGAGCGCTGGAAC 1620  
Db 1561 GCGCTGAGAGAGCGGGCAACGGCGCAACCTTGTGCGGAGCGCGCGAGCGCTGGAAC 1620  
QY 1621 GCCCTGAAGAGCTGTACGAAGACCGCGGAGAGCTTGTGCGGAGCGCGCGAGCGCTGGAAG 1680  
Db 1621 GCCCTGAAGAGCTGTACGAAGACCGCGGAGAGCTTGTGCGGAGCGCGCGAGCGCTGGAAG 1680  
QY 1681 AAGCACTTCGGAAGACACAGCAGCGCTGACGCTCCGGCAGCAAGGGCGGAAGAGCGG 1740  
Db 1681 AAGCACTTCGGAAGACACAGCAGCGCTGACGCTCCGGCAGCAAGGGCGGAAGAGCGG 1740  
QY 1741 CTTGCCGAATTTGAAGCGCGCGCAAGTTCCTTGAACCAATGTTCCCGGAA 1800  
Db 1741 CTTGCCGAATTTGAAGCGCGCGCAAGTTCCTTGAACCAATGTTCCCGGAA 1800  
QY 1801 GACGCGAGCTGACCGGAGCGGCGCTAAGTGTGTGGGGCGCGCTCAGTAGACGAC 1860  
Db 1801 GACGCGAGCTGACCGGAGCGGCGCTAAGTGTGTGGGGCGCGCTCAGTAGACGAC 1860  
QY 1861 AAGCGGTGTTGCTGCGGCTTTCGTAAGACAAGATGTTGACGAAGTGCAGTACGGGC 1920  
Db 1861 AAGCGGTGTTGCTGCGGCTTTCGTAAGACAAGATGTTGACGAAGTGCAGTACGGGC 1920  
QY 1921 AGGGGCGAGGAAGCGCCATCGAGAAGCGCGCTTCGATCAGTGGGGAAGCGCGGAC 1980  
Db 1921 AGGGGCGAGGAAGCGCCATCGAGAAGCGCGCTTCGATCAGTGGGGAAGCGCGGAC 1980  
QY 1981 GACGACGAGGAAGAGCAGCGCCCAAGAGCGGACAGGAAGAGCTAGCGGCTAG 2031  
Db 1981 GACGACGAGGAAGAGCAGCGCCCAAGAGCGGACAGGAAGAGCTAGCGGCTAG 2031

RESULT 3  
US-09-940-550A-11  
; Sequence 11, Application US/09940550A  
; Publication No. US20030033635A1

```

: GENERAL INFORMATION:
: APPLICANT: MANKIN, LUKE
: APPLICANT: MCKERSIE, BRYAN
: TITLE OF INVENTION: SELF-EXCISING POLYNUCLEOTIDES AND USES THEREOF
: FILE REFERENCE: 16313-0055
: CURRENT APPLICATION NUMBER: US/09/940,550A
: PRIOR APPLICATION NUMBER: 2001-08-27
: PRIOR FILING DATE: 2000-08-25
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO 11
: LENGTH: 9901
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: PBPS EW051
: NAME/KEY: modified_base
: LOCATION: (1984)
: OTHER INFORMATION: a, t, c, g, other or unknown
: NAME/KEY: modified_base
: LOCATION: (1986)
: OTHER INFORMATION: a, t, c, g, other or unknown
: NAME/KEY: modified_base
: LOCATION: (5231)
: OTHER INFORMATION: a, t, c, g, other or unknown
: NAME/KEY: modified_base
: LOCATION: (5233)
: OTHER INFORMATION: a, t, c, g, other or unknown
: NAME/KEY: modified_base
: LOCATION: (8478)
: OTHER INFORMATION: a, t, c, g, other or unknown
: NAME/KEY: modified_base
: LOCATION: (8480)
: OTHER INFORMATION: a, t, c, g, other or unknown
: US-09-940-550A-11

```

Query Match 98.4%; Score 1998; DB 9; Length 9901;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 2021; Conservative 0; Mismatches 10; Indels 7; Gaps 1;

```

QY 1 ATGCACAAAGGGGTGTGACCGGGGTGATACGTAAGTTCTGCTTCTACCTTTGATATA 60
Db 4984 ATGCACAAAGGGGTGTGACCGGGGTGATACGTAAGTTCTGCTTCTACCTTTGATATA 5043
QY 61 TATATAATATATATATATATATATATATATATATATATATATATATATATATATAT 120
Db 5044 TATATAATATATATATATATATATATATATATATATATATATATATATATATATAT 5103
QY 121 AAAGATGTAGTATATAGCAATGCTTTCTGTAGTTATAGTGTATATATATATATAT 180
Db 5104 AAAGATGTAGTATATAGCAATGCTTTCTGTAGTTATAGTGTATATATATATATATAT 5163
QY 181 ATAACTTTTCTAATATATATAGCAAAATTTGTGTATGTGACAGGTACGGGGGTCTTACGAC 240
Db 5164 ATAACTTTTCTAATATATATAGCAAAATTTGTGTATGTGACAGGTACGGGGGTCTTACGAC 5223
QY 241 CGTC-----AGTCGGCGGCGGAGAGATAGCAGTGCAGCAAGCCACGACACAGCG 293
Db 5224 CGTCGRNTNDAGTCGGCGGCGGAGAGATAGCAGTGCAGCAAGCCACGACACAGCG 5283
QY 294 TAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 353
Db 5284 TAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5343
QY 354 GTTCAGGTTGTCGGGCAATTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 413
Db 5344 GTTCAGGTTGTCGGGCAATTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5403
QY 414 GCGCCCGAGAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 473
Db 5404 GCGCCCGAGAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5463

```

```

QY 474 TGTCTATGACGTGTCGGCTTCTCGCCCTGAAGGTATGAGCGGATTCGATGTC 533
Db 5464 TGTCTATGACGTGTCGGCTTCTCGCCCTGAAGGTATGAGCGGATTCGATGTC 5523
QY 534 GGAATTGTCGGCCTGGGCGTGAAGATTTGTTCCACTCAGAGAGAGAGAGAGAGAGAG 593
Db 5524 GGAATTGTCGGCCTGGGCGTGAAGATTTGTTCCACTCAGAGAGAGAGAGAGAGAGAG 5583
QY 594 AAACGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 653
Db 5584 AAACGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5643
QY 654 GCTGAAGTGGCGGAGAGATTTCTGACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 713
Db 5644 GCTGAAGTGGCGGAGAGATTTCTGACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5703
QY 714 CGCGGGAAGGCGCTTACGGCTTCAGAGCTTGTTCGAGAGAGAGAGAGAGAGAGAGAG 773
Db 5704 CGCGGGAAGGCGCTTACGGCTTCAGAGCTTGTTCGAGAGAGAGAGAGAGAGAGAGAG 5763
QY 774 CGCGGGAATGTCATATGTCGTATCAACAAGTTAGCGCACTGACCACTCCCTTACCGG 833
Db 5764 CGCGGGAATGTCATATGTCGTATCAACAAGTTAGCGCACTGACCACTCCCTTACCGG 5823
QY 834 ACCCTTCGAGTTGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 893
Db 5824 ACCCTTCGAGTTGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5883
QY 894 ACACCTTCCTTCAAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 953
Db 5884 ACACCTTCCTTCAAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5943
QY 954 TTGTAAGCGATGAGAGCGTGAAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1013
Db 5944 TTGTAAGCGATGAGAGCGTGAAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6003
QY 1014 CGCTTCAAGCGCTGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1073
Db 6004 CGCTTCAAGCGCTGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6063
QY 1074 GGGCTTCGCGCTGAGGTGATCTACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1133
Db 6064 GGGCTTCGCGCTGAGGTGATCTACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 6123
QY 1134 TGAGGTTACCGCATTCAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1193
Db 6124 TGAGGTTACCGCATTCAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6183
QY 1194 ACCGATCATGAGAGCGGCTGAGTGTATGAGCTTCAGGCGTGTGAGAGAGAGAGAGAG 1253
Db 6184 ACCGATCATGAGAGCGGCTGAGTGTATGAGCTTCAGGCGTGTGAGAGAGAGAGAGAG 6243
QY 1254 CGGCAAGGGGCTTCCCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1313
Db 6244 CGGCAAGGGGCTTCCCGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6303
QY 1314 GTGTGGCGCGGTATGACTTCGAAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1373
Db 6304 GTGTGGCGCGGTATGACTTCGAAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6363
QY 1374 CGGTGCGGGAAGGTGCTGACCGGCTGCGCACTGGGAGAGAGAGAGAGAGAGAGAGAG 1433
Db 6364 CGGTGCGGGAAGGTGCTGACCGGCTGCGCACTGGGAGAGAGAGAGAGAGAGAGAGAGAG 6423
QY 1434 CAGCATGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1493
Db 6424 CAGCATGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6483
QY 1494 CGAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1553
Db 6484 CGAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6543

```

QY	1554	CACTGAGGCGCCTGAGAGACGCGGGCGAACGGGGCGAACCTTGTGTGGGAGCGCGCCGCAACC	1613
Db	6544	CACTGAGGCGCCTGAGAGAGCGGGCGAACGGGGCGAACCTTGTGTGGGAGCGCGCCGCAACC	6603
QY	1614	CCTGAACGCCCCCTTGAAGAGCTGTACGAAGACCGCGCGCGAGGAGCTTACGACGGACCCGT	1673
Db	6604	CCTGAACGCCCCCTTGAAGAGCTGTACGAAGACCGCGCGCGAGGAGCTTACGACGGACCCGT	6663
QY	1674	TGGCAGGAAGCACTTCCGGAAGCAACAGGCGAGCGCTGACGCTCCGGCAGCAAGGGCGCGA	1733
Db	6664	TGGCAGGAAGCACTTCCGGAAGCAACAGGCGAGCGCTGACGCTCCGGCAGCAAGGGCGCGA	6723
QY	1734	AGAGCGGCTTGCAGCACTTGAAGCCCGCGAAGCCCCGGAAGTTGCCCTTGACCAATGGTT	1793
Db	6724	AGAGCGGCTTGCAGCACTTGAAGCCCGCGAAGCCCCGGAAGCTTCCCTTGACCAATGGTT	6783
QY	1794	CCCCGAAGACGCGGAGCTGACCCCGACCGGCCCTAAGTCGTGGTGGGGCGCGGCTCAGT	1853
Db	6784	CCCCGAAGACGCGGAGCTGACCCCGACCGGCCCTAAGTCGTGGTGGGGCGCGGCTCAGT	6843
QY	1854	AGACGACAAGCGCGTGTCTCGGGGCTCTTCTGTAGACAGATCGTTGTCACGAAGTCGAC	1913
Db	6844	AGACGACAAGCGCGTGTCTCGGGGCTCTTCTGTAGACAGATCGTTGTCACGAAGTCGAC	6903
QY	1914	TACGGGCGAGGGGCGAGGGGAACGCCCATCGAGAAGCGCGCTTCATCAGTGGGGCGAAGCC	1973
Db	6904	TACGGGCGAGGGGCGAGGGGAACGCCCATCGAGAAGCGCGCTTCATCAGTGGGGCGAAGCC	6963
QY	1974	GCCGACCGACGACGACGAAGACGACGCCCGAGGACGGGCACGGAAGACGTAGCGGCGTAG	2031
Db	6964	GCCGACCGACGACGACGAAGACGACGCCCGAGGACGGGCACGGAAGACGTAGCGGCGTAG	7021

## RESULT 4

```

US-09-788-297-20
; Sequence 20, Application US/09788297
; Patent No. US20020094516A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Calos, Michele P.
; APPLICANT: Scilmentl, Christopher R.
; TITLE OF INVENTION: ALTERED RECOMBINASES FOR GENOME MODIFICATION
; FILE REFERENCE: 8400-0011
; CURRENT APPLICATION NUMBER: US/09/788, 297
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
;
; LENGTH: 1842
;
; TYPE: DNA
;
; ORGANISM: Phage phiC31
US-09-788-297-20

```

Query Match	88.48;	Score 1795;	DB 10;	Length 1842;
Best Local Similarity	99.28;	Pred. No. 0;		
Matches 1804; Conservative	0;	Mismatches 15;	Indels 0;	Gaps 0;

QY	213	GA	GTG	CAG	GTAC	GGCGGGT	GCTTAC	GACCC	GTCA	GTGCGCGG	CAGCGG	GAATAG	CATG	272																												
	11	11	11	11	11	11	11	11	11	11	11	11	11																													
Db	24	GGT	GAC	ACG	TAC	CGGGT	GCTTAC	GACCC	GTCA	GTGCGCGG	GAATAG	CATG	CGC	83																												
QY	273	AGCA	AGCC	CAG	CGAC	CACAG	CGT	AGCG	CCCA	CGAAG	CGCGG	CCGAC	CTT	CAG	CGCGA	332																										
	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11																											
Db	84	AGCA	AGCC	CAG	CGAC	CACAG	CGT	AGCG	CCCA	CGAAG	CGCGG	CCGAC	CTT	CAG	CGCGA	143																										
QY	333	AGT	C	GAG	CGCG	CGGGGG	CCGGT	C	T	CAG	GTTC	G	T	C	G	G	C	A	T	T	C	A	G	G	C	G	G	C	A	C	392											
	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11												
Db	144	AGT	C	GAG	CGCG	CGGGGG	CCGGT	C	T	CAG	GTTC	G	T	C	G	G	C	A	T	T	C	A	G	G	C	G	C	A	C	203												
QY	393	GTC	G	G	C	G	CTT	C	G	G	A	C	G	G	G	C	G	G	A	G	C	G	C	C	C	G	A	G	T	T	C	G	A	A	T	G	C	G	C	452		
	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11						
Db	204	GTC	G	G	C	G	CTT	C	G	G	A	C	G	G	G	C	G	G	A	G	C	G	C	C	C	G	A	G	T	T	C	G	A	A	T	G	C	G	C	263		
QY	453	CGG	G	CG	G	C	T	C	A	C	A	T	G	A	T	G	A	T	G	A	C	G	T	G	C	G	C	T	T	C	G	C	C	T	G	A	G	T	C	A	T	512

Db	264	CGGGCGGCTCACATGATCATTTCTATGACGTGTGGCGCTTCGCGCCTGAAGGTCAT	323
QY	513	GGACCGGATTCGGATTGTCTCGGAATTGCTCGCCCTGGGCGTGACGATTGTTCCACTCA	572
Db	324	GGACCGGATTCGGATTGTCTCGGAATTGCTCGCCCTGGGCGTGACGATTGTTCCACTCA	383
QY	573	GGAAAGCGTCTTCGGGACGGGAACGTCATGGACCTGATTCACTGATTATGCGGCTCGA	632
Db	384	GGAAAGCGTCTTCGGGACGGGAACGTCATGGACCTGATTCACTGATTATGCGGCTCGA	443
QY	633	CGCGTCGCACAAAGAATCTTCGCTGAAGTTCGGCGGAAGATTCTGCACAGAAACCTTCA	692
Db	444	CGCGTCGCACAAAGAATCTTCGCTGAAGTTCGGCGGAAGATTCTGCACAGAAACCTTCA	503
QY	693	GGCGGAATTGGGCGGGTACGTCGGCGGGGAAGGCGCCTTACGGCTTCGAGCTTGTTCGGA	752
Db	504	GGCGGAATTGGGCGGGTACGTCGGCGGGGAAGGCGCCTTACGGCTTCGAGCTTGTTCGGA	563
QY	753	GACGAAGGAGATCACGCGGCAACGGCCGAATGTGTCATGTGCTCATCAACAAGTTAGCGCA	812
Db	564	GACGAAGGAGATCACGCGGCAACGGCCGAATGTGTCATGTGCTCATCAACAAGTTAGCGCA	623
QY	813	CTGCACCACTCCCTTACCGGACCCCTTCGAGTTCGAGCCCCGACGTAATCCGGTGGTGGTG	872
Db	624	CTGCACCACTCCCTTACCGGACCCCTTCGAGTTCGAGCCCCGACGTAATCCGGTGGTGGTG	683
QY	873	GCGTGAGATCAAGACGCACAAACACTTCCTCTCAAGCCGGCGAGTCAAGCCGCCATTCA	932
Db	684	GCGTGAGATCAAGACGCACAAACACTTCCTCTCAAGCCGGCGAGTCAAGCCGCCATTCA	743
QY	933	CCCGGCGACATCACGGGGCTTTGTAAGCGCATGGACGCTGACGCCGTGCGACCCGGGG	992
Db	744	CCCGGCGACATCACGGGGCTTTGTAAGCGCATGGACGCTGACGCCGTGCGACCCGGGG	803
QY	993	CGAGACGATTGGGAAGAAGACCCTTCAAGCGCCTGGGACCCGGCAACCGTTATGGCAAT	1052
Db	804	CGAGACGATTGGGAAGAAGACCCTTCAAGCGCCTGGGACCCGGCAACCGTTATGGCAAT	863
QY	1053	CCTTCCGGGACCCCGCTATTGCGGGCTTCGCCGCTGAGGTGATCTACAAGAAGACCGGA	1112
Db	864	CCTTCCGGGACCCCGCTATTGCGGGCTTCGCCGCTGAGGTGATCTACAAGAAGACCGGA	923
QY	1113	CGGACGCGCGGACCGACGGAAGATTGAGGGTTACCGCATTCAGCGCGAACCAGATCAGCCTCG	1172
Db	924	CGGACGCGCGGACCGACGGAAGATTGAGGGTTACCGCATTCAGCGCGAACCAGATCAGCCTCG	983
QY	1173	GCCGCTGAGCTTGATTGCGGACCGATCATCGAGCCCGCTGAGTGTATGAGCTTCAGGC	1232
Db	984	GCCGCTGAGCTTGATTGCGGACCGATCATCGAGCCCGCTGAGTGTATGAGCTTCAGGC	1043
QY	1233	GTTGTTGACGGCAGGGGGCGGCAAGGGGCTTTCGCCGGGGCAAGCCATTCTGTCCGC	1292
Db	1044	GTTGTTGACGGCAGGGGGCGGCAAGGGGCTTTCGCCGGGGCAAGCCATTCTGTCCGC	1103
QY	1293	CATGACACAAGCTGTACTCGGAGTGTGGCCCGCTCATGACTTTCGAAGCGCGGGAAGAATC	1352
Db	1104	CATGACACAAGCTGTACTCGGAGTGTGGCCCGCTCATGACTTTCGAAGCGCGGGAAGAATC	1163
QY	1353	GATCAAGGACTCTTACCGCTGCCGTGCGGGAAGGTGGTGCACCCGTCGCCACCTGGGCA	1412
Db	1164	GATCAAGGACTCTTACCGCTGCCGTGCGGGAAGGTGGTGCACCCGTCGCCACCTGGGCA	1223
QY	1413	GCACGAAGGCGCTGCAACGTCACATGGCGGCACTGCACAAGTTCGTTGCGGAACGCAT	1472
Db	1224	GCACGAAGGCGCTGCAACGTCACATGGCGGCACTGCACAAGTTCGTTGCGGAACGCAT	1283
QY	1473	CTTCAACAAGATCAGGCACGCCGAAGCGCAGCAAGAGACGTTGGCGCTTCTGTGGAAAC	1532
Db	1284	CTTCAACAAGATCAGGCACGCCGAAGCGCAGCAAGAGACGTTGGCGCTTCTGTGGAAAC	1343
QY	1533	CGCCGACGCTTCGGCAAGCTCACTGAGCGCGCTGAGAAGAGCGGCGAAGCGGCAACCT	1592

```

Db 1344 CGCCGACGCTTCGGCAAGCTCTACTGAGCGCCCTGAGAGACGGCGGAACGGCGCAACCT 1403
QY 1593 TGTTCGGAGCGCGCCGACGCGCCCTGAGACGCCCTTGAAGAGCTGTACGAAGACGGCGCGC 1652
Db 1404 TGTTCGGAGCGCGCCGACGCGCCCTGAGACGCCCTTGAAGAGCTGTACGAAGACGGCGCGC 1463
QY 1653 AGGAGCTTACGACGCGACCCGTTGGCAGGAAGACACTTCCGGAAGCAACAGGACGGCTGAC 1712
Db 1464 AGGCGGTACGACGCGACCCGTTGGCAGGAAGCACTTCCGGAAGCAACAGGACGGCTGAC 1523
QY 1713 GCTCCGGCAGCAAGGGCGGGAAGAGCGGCTTGGCGAACTTGAAGCCGCGGAAAGCCCGAA 1772
Db 1524 GCTCCGGCAGCAAGGGCGGGAAGAGCGGCTTGGCGAACTTGAAGCCGCGGAAAGCCCGAA 1583
QY 1773 GTTCCCTTGAACCAATGTTCCTCCGGAAGACGCGCGCTGACCCGACGGCGCTAAGTC 1832
Db 1584 GCTTCCCTTGAACCAATGTTCCTCCGGAAGACGCGCGCTGACCCGACGGCGCTAAGTC 1643
QY 1833 GTGTTGGGGCGCGCGTCACTAGACGACGACGCGCTGCTCGGCTCTTCTAGACAA 1892
Db 1644 GTGTTGGGGCGCGCGTCACTAGACGACGACGCGCTGCTCGGCTCTTCTAGACAA 1703
QY 1893 GATGTTGTACAGAGTCACTAGACGAGGGGAGGGGAGGGAAGCGCCCATCGAAGCGCGC 1952
Db 1704 GATGTTGTACAGAGTCACTAGACGAGGGGAGGGGAGGGAAGCGCCCATCGAAGCGCGC 1763
QY 1953 TTGCATCAGCTGGGCGAAGCGCGCGCGACGACGACGACGAGAGAGAGCGCCCGAGGCGAC 2012
Db 1764 TTGCATCAGCTGGGCGAAGCGCGCGCGACGACGACGAGAGAGAGCGCCCGAGGCGAC 1823
QY 2013 GGAAGACGTAGCGGCGTAG 2031
Db 1824 GGAAGACGTAGCGGCGTAG 1842

```

RESULT 5  
US-09-788-297-26

```

: Sequence 26, Application US/09788297
: Patent No. US20020094516A1
: GENERAL INFORMATION:
: APPLICANT: Calos, Michele P.
: APPLICANT: Scilment, Christopher R.
: TITLE OF INVENTION: ALTERED RECOMBINASES FOR GENOME MODIFICATION
: FILE REFERENCE: 8400-0011
: CURRENT APPLICATION NUMBER: US/09/788, 297
: CURRENT FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 26
: LENGTH: 1842
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: altered
: US-09-788-297-26

```

Query Match 88.1%; Score 1790.2; DB 10; Length 1842;  
Best Local Similarity 99.0%; Pred. No. 0;  
Matches 1801; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```

QY 213 GATGTGACAGTACGCGGCTTACGACCGTCACTGCGGCGAGCGGAGAGATAGCAAGTGC 272
Db 24 GGTGACACGTACGCGGCTTACGACCGTCACTGCGGCGAGCGGAGAAATTCAGAGCGC 83
QY 273 AGCAAGCCGACGACACAGAGCTAGCGGCCAACAGAGACAGAGCGCGCGACCTTCAGCGCGA 332
Db 84 AGCAAGCCGACGACACAGAGCTAGCGGCCAACAGAGACAGAGCGCGCGACCTTCAGCGCGA 143
QY 333 AGTCAGCGCGAGCGGGGCGGCTTCAAGTTCGTCGGGCAATTCAGGAGAGCGCGGCGAC 392
Db 144 AGTCAGCGCGAGCGGGGCGGCTTCAAGTTCGTCGGGCAATTCAGGAGAGCGCGGCGAC 203

```

```

QY 393 GTCCGCGTTCGGAGCGCGGAGCGCCCGGAGTTCCGAAAGCATCTCTGAACGAATGCCCGC 452
Db 204 GTCCGCGTTCGGAGCGCGGAGCGCCCGGAGTTCCGAAAGCATCTCTGAACGAATGCCCGC 263
QY 453 CGGGCGGCTCAACATGATCATTTGTCTATGAGCTGTCCGCTTCTCCGCCCTGAAGTCTAT 512
Db 264 CGGGCGGCTCAACATGATCATTTGTCTATGAGCTGTCCGCTTCTCCGCCCTGAAGTCTAT 323
QY 513 GACGCGATTTCCGATTTCTCGGAATTTGCTCGCCCTGGGCGTGAAGATTTGTTCCACTCA 572
Db 324 GACGCGATTTCCGATTTCTCGGAATTTGCTCGCCCTGGGCGTGAAGATTTGTTCCACTCA 383
QY 573 GGAAGGCGTCTTCCGCGAGGGAAGCGTCACTGACCTGATTCACCTGATTTATGCGGCTCGA 632
Db 384 GGAAGGCGTCTTCCGCGAGGGAAGCGTCACTGACCTGATTCACCTGATTTATGCGGCTCGA 443
QY 633 CGCGTCCACAAAGATCTTCTGTAAGTCCGCGAAGATTTCTGACACGAAAGACCTTCA 692
Db 444 CGCGTCCACAAAGATCTTCTGTAAGTCCGCGAAGATTTCTGACACGAAAGACCTTCA 503
QY 693 GCGGCAATTGGCGGCTACGTCGCGGGAAGCGCGCTTACGGCTTCGAGCTTGTTCGGA 752
Db 504 GCGGCAATTGGCGGCTACGTCGCGGGAAGCGCGCTTACGGCTTCGAGCTTGTTCGGA 563
QY 753 GACGAAGGAGATCAAGCGCAACGCGCGCAATGTCAATGTCTCATCAACAAGTTAGCGCA 812
Db 564 GACGAAGGAGATCAAGCGCAACGCGCGCAATGTCAATGTCTCATCAACAAGTTAGCGCA 623
QY 813 CTCGACCACTCCCTTACCAGCCCTTCGAGTTCGAGCCCGACGTAATCCGCTGTGTG 872
Db 624 CTCGACCACTCCCTTACCAGCCCTTCGAGTTCGAGCCCGACGTAATCCGCTGTGTG 683
QY 873 GCGTGAGATCAAGACGACACAACACCTTCCCTCAAGCCGCGGAGTCAAGCCCGCATTCA 932
Db 684 GCGTGAGATCAAGACGACACAACACCTTCCCTCAAGCCGCGGAGTCAAGCCCGCATTCA 743
QY 933 CCGGCGGACGATCACGGGCTTTGTAAGCGCATGAGCCGTGACCGCGCGGCGG 992
Db 744 CCGGCGGACGATCACGGGCTTTGTAAGCGCATGAGCCGTGACCGCGCGGCGG 803
QY 993 CGAGACGATTTGGGAAGAGACCGCTTCAAGCGCGCTGGACCGCGCAACCGTATGCGAAT 1052
Db 804 CGAGACGATTTGGGAAGAGACCGCTTCAAGCGCGCTGGACCGCGCAACCGTATGCGAAT 863
QY 1053 CCTTCGGGACCGCGTATTGCGGCGCTTCGCGCTGAGGTGATCTACAAGAGAACCGGA 1112
Db 864 CCTTCGGGACCGCGTATTGCGGCGCTTCGCGCTGAGGTGATCTACAAGAGAACCGGA 923
QY 1113 CGGACCGCGACACGAGAGATGAGGGTTACCGCATTCAGCGCGACCGCATCGCTCCG 1172
Db 924 CGGACCGCGACACGAGAGATGAGGGTTACCGCATTCAGCGCGACCGCATCGCTCCG 983
QY 1173 GCGGTCGAGCTTGATTGCGGACCGATCATGAGCCCGCTGAGTGTATGAGCTTCAGGC 1232
Db 984 GCGGTCGAGCTTGATTGCGGACCGATCATGAGCCCGCTGAGTGTATGAGCTTCAGGC 1043
QY 1233 GTGTTGACGCGGAGGGGCGGCAAGGGGCTTCCCGGGGCAAGCATTTCTGCCG 1292
Db 1044 GTGTTGACGCGGAGGGGCGGCAAGGGGCTTCCCGGGGCAAGCATTTCTGCCG 1103
QY 1293 CATGCAAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAAGCGCGGGAAGAATC 1352
Db 1104 CATGCAAGCTGTACTGCGAGTGTGGCGCGCTCATGACTTCGAAGCGCGGGAAGAATC 1163
QY 1353 GATCAAGGACTTTACCGCTCCGCTGCGCGGAGAGGTGTGACCCCGTCCGACCTGGGCA 1412
Db 1164 GATCAAGGACTTTACCGCTCCGCTGCGCGGAGAGGTGTGACCCCGTCCGACCTGGGCA 1223
QY 1413 GCACGAAGGACGTCGACGTCAGCATGGCGCACTGACAAAGTTGTCGGAAGCAT 1472
Db 1224 GCACGAAGGACGTCGACGTCAGCATGGCGCACTGACAAAGTTGTCGGAAGCAT 1283
QY 1473 CTTCAACAAGATCAGGACGCGCGAAGGCGAGCAAGAGACGTTGGCGCTTCTGTGGAAGC 1532

```



Db	1284	CTTCAACAAGATCAGGCGACGCCGAAGCGCAGCAAGACGTTGGCGCTTCTGTGGGAAGC	1343
QY	1533	CGCCCGACGCTTCGGCAAGCTCACTGAGGCGCCTGAGAAGAGCGGCGAACCT	1592
Db	1344	CGCCCGACGCTTCGGCAAGCTCACTGAGGCGCCTGAGAAGAGCGGCGAACCT	1403
QY	1593	TGTTGGGAGGCGCGCCGACGCCCTGAACGCCCTTTGAAGAGCTGTACGAAGACCGCGCGC	1652
Db	1404	TGTTGGGAGGCGCGCCGACGCCCTGAACGCCCTTTGAAGAGCTGTACGAAGACCGCGCGC	1463
QY	1653	AGGAGCTTACGACGAGACCCGTTGGCAGGAAGCACTCCGGAACACAGCAGCGCTGAC	1712
Db	1464	AGGCGCGTACGACGAGACCCGTTGGCAGGAAGCACTCCGGAACACAGCAGCGCTGAC	1523
QY	1713	GCTCCGGCAGCAAGGGGCGGAGAGAGCGGCTTGCCGAACCTGAAGCCGCCGAAGCCCCGA	1772
Db	1524	GCTCCGGCAGCAAGGGGCGGAGAGAGCGGCTTGCCGAACCTGAAGCCGCCGAAGCCCCGA	1583
QY	1773	GTTGCCCTTGACCAATGTTCCCGGAAGACGCCGACGCTGACCCGACCGGCCCTTAAGTC	1832
Db	1584	GCTTCCCTTGACCAATGTTCCCGGAAGACGCCGACGCTGACCCGACCGGCCCTTAAGTC	1643
QY	1833	GTTGGTGGGGGCGCGCTCACTAGACGACAAAGCGCGTTCGTGGGCTCTTCGTAGACAA	1892
Db	1644	GTTGGTGGGGGCGCGCTCACTAGACGACAAAGCGCATGTTCTTGGGCTCTTCGTAGACAA	1703
QY	1893	GATCGTTGTACGAAGTCACTACGGGCGAGGGGCGAGGGAACGCCATCGAGAAACGCGC	1952
Db	1704	GATCGTTGTACGAAGTCACTACGGGCGAGGGGCGAGGGAACGCCCATCGAGAAACGCGC	1763
QY	1953	TTTCGATCAGTGGGCGAAGCCCGACCGACGACGACGAAGACGACGCCACGACGGCAC	2012
Db	1764	TTTCGATCAGTGGGCGAAGCCCGACCGACGACGACGACGAAGACGACGCCACGACGGCAC	1823
QY	2013	GGAAGACGTAGCGGCGTAG 2031	
Db	1824	GCAAGACGTAGCGGCGTAG 1842	

## RESULT 6

```

US-09-788-297-27
; Sequence 27, Application US/09788297
; Patent No. US20020094516a1
; GENERAL INFORMATION:
; APPLICANT: Calos, Michele P.
; APPLICANT: Scilmentl, Christopher R.
; TITLE OF INVENTION: ALTERED RECOMBINASES FOR GENOME MODIFICATION
; FILE REFERENCE: 8400-0011
; CURRENT APPLICATION NUMBER: US/09/788,297
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 1839
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: altered
US-09-788-297-27

```

Query Match	87.9%	Score 1785.6;	DB 10;	Length 1839;
Best Local Similarity	99.0%;	Pred. No. 0;		
Matches 1797; Conservative	0;	Mismatches 19;	Indels 0;	Gaps 0;

[illegible]

QY	333	AGTCAGCGCGACGGGGCCGGTTCAGGTTCTGTCGGGCATTTTCAGCGAAGCGCCGGGCAC	392
	144	AGTCGAGCGCGACGGGGCCGGTTCAGATTTCGTGGGCATTTTCAGCGAAGCGCCGGGCAC	203
QY	393	GTCGGCGTTCCGGGACGGCGGAGCCCGGAGTTGAAAGCATCTGAACGAATGCCGCGC	452
	204	GTCGGCGTTCCGGGACGGCGGAGCCCGGAGTTGAAAGCATCTGAACGAATGCCGCGC	263
QY	453	CGGCGGCTCAACATGATCATTTGTTCTATGACGTGTCCGCTTCTCGCGCTGAAGTTCAT	512
	264	CGGCGGCTCAACATGATCATTTGTTCTATGACGTGTCCGCTTCTCGCGCTGAAGTTCAT	323
QY	513	GGACCGGATTCGGATTGTCTCGGAATTTGCTCGCCCTGGGCGTGACGATTGTTCCACTCA	572
	324	GGACCGGATTCGGATTGTCTCGGAATTTGCTCGCCCTGGGCGTGACGATTGTTCCACTCA	383
QY	573	GGAAGCGCTTTCGGGACGGAACGTATGAGCCGTATTCACCTGATTATGCGGCTCGA	632
	384	GGAAGCGCTTTCGGGACGGAACGTATGAGCCGTATTCACCTGATTATGCGGCTCGA	443
QY	633	CGCGTCGCACAAGAATCTTCGCTGAAGTTCGGCGAAGATTCTGCACACGAAGAACCTTCA	692
	444	CGCGTCGCACAAGAATCTTCGCTGAAGTTCGGCGAAGATTCTGCACACGAAGAACCTTCA	503
QY	693	GCGCGAATTGGGCGGGGTACGTCGGCGGGGAAGGCGCCTTACGCGCTTCGAGCTTGTTCGGA	752
	504	GCGCGAATTGGGCGGGGTACGTCGGCGGGGAAGGCGCCTTACGCGCTTCGAGCTTGTTCGGA	563
QY	753	GACGAAGAGATCACGCGCAACGGCCGAATGTTCAATGTCTCATCAACAAGTTACGCA	812
	564	GACGAAGAGATCACGCGCAACGGCCGAATGTTCAATGTCTCATCAACAAGTTACGCA	623
QY	813	CTGACCACTCCCTTACCAGCCCTTCGAGTTCGAGCCCGACGTAATCCGGTGTGCTG	872
	624	CTGACCACTCCCTTACCAGCCCTTCGAGTTCGAGCCCGACGTAATCCGGTGTGCTG	683
QY	873	GCGTAGATCAAGACGCACAAACACTTCCCTTCAAGCCGGCAGTCAAGCCGCCATTC	932
	684	GCGTAGATCAAGACGCACAAACACTTCCCTTCAAGCCGGCAGTCAAGCCGCCATTC	743
QY	933	CCCCGGCAGCATCACGGGGCTTTGTAAGCGCATGGACGCTGACGCCGTGCCGACCCGGG	992
	744	CCCCGGCAGCATCACGGGGCTTTGTAAGCGCATGGACGCTGACGCCGTGCCGACCCGGG	803
QY	993	CGAGACGATTGGGAAGAGACCCTTCAAGCGCCTGGGACCCGGCAACCGTTATGCAAT	1052
	804	CGAGACGATTGGGAAGAGACCCTTCAAGCGCCTGGGACCCGGCAACCGTTATGCAAT	863
QY	1053	CCTTCGGGACCCCGCGTATTGCGGGCTTCGCCGCTGAGGTGATCTACAAGAAGACC	1112
	864	CCTTCGGGACCCCGCGTATTGCGGGCTTCGCCGCTGAGGTGATCTACAAGAAGACC	923
QY	1113	CGGCACGCCGACCAACGAAGATTGAGGGTTACCGCATTCACGCGCAACCCGATCACGCTCG	1172
	924	CGGCACGCCGACCAACGAAGATTGAGGGTTACCGCATTCACGCGCAACCCGATCACGCTCG	983
QY	1173	GCCGGTGCAGCTTGATTGGCGGACCATCATGAGCCCGCTGAGTGATGAGCTTCAAGC	1232
	984	GCCGGTGCAGCTTGATTGGCGGACCATCATGAGCCCGCTGAGTGATGAGCTTCAAGC	1043
QY	1233	GTCGTTGCACGCGACGGGGCGCGCAAGGGCTTTCGCCGGGGCAAGCCATTCTGTCCGC	1292
	1044	GTCGTTGCACGCGACGGGGCGCGCAAGGGCTTTCGCCGGGGCAAGCCATTCTGTCCGC	1103
QY	1293	CATGACAAGCTGTACTGCGAGTGTGCGCCGCTCATGACTTCGAAGCGCGGGAAGATC	1352
	1104	CATGAGGCAAGCTGTACTGCGAGTGTGCGCCGCTCATGACTTCGAAGCGCGGGAAGATC	1163
QY	1353	GATCAAGGACTCTTACCGCTGCCGTGCGCGGAAGGTGTTGACACCCGTCCGACACTGGCA	1412
	1164	GATCAAGGACTCTTACCGCTGCCGTGCGCGGAAGGTGTTGACACCCGTCCGACACTGGCA	1223



QY	1413	GCACGAAGGCACGTTGCAACGTCAGCATGGCGGCGCACTCGACCAAGTTCGTTGCCGAACGCAT	1472
Db	1224	GCACGAAGGCACGTTGCAACGTCAGCATGGCGGCGCACTCGACCAAGTTCGTTGCCGAACGCAT	1283
QY	1473	CTTCAACAAGATCAGGCACGCCGGAAGCCGACGAAGAGAGACGTTGGCGCTTCTGTGGGAAGC	1532
Db	1284	CTTCAACAAGATCAGGCACGCCGGAAGCCGACGAAGAGAGACGTTGGCGCTTCTGTGGGAAGC	1343
QY	1533	CGCCCGACCGTTCGGCAAGCTCACTGAGGCGCCTGAGAAGAGCGGCGGAACGGCGCAACT	1592
Db	1344	CGCCCGACCGTTCGGCAAGCTCACTGAGGCGCCTGAGAAGAGCGGCGGAACGGCGCAACT	1403
QY	1593	TGTTGCGGAGCGCGCCGACGCCCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGC	1652
Db	1404	TGTTGCGGAGCGCGCCGACGCCCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGC	1463
QY	1653	AGGAGCTTACGACGAGACCCGTTGGCAGGAAGCAGCTTCCGGAAGCAACAGGACAGCGCTGAC	1712
Db	1464	AGGCGCGTACGACGAGACCCGTTGGCAGGAAGCAGCTTCCGGAAGCAACAGGACAGCGCTGAC	1523
QY	1713	GCTCCGGCAGCAAGGGGCGGAAGAGCGGCTTGCCGAAGCTTGAAGCCGCCGAAGCCCGAA	1772
Db	1524	GCTCCGGCAGCAAGGGGCGGAAGAGCGGCTTGCCGAAGCTTGAAGCCGCCGAAGCCCGAA	1583
QY	1773	GTTGCCCCCTTGACCAATGTTCCCGGAAGACGCCGAGCTGACCCGACCGGCCCTAAGTC	1832
Db	1584	GCTTCCCTTGACCAATGTTCCCGGAAGACGCCGAGCTGACCCGACCGGCCCTAAGTC	1643
QY	1833	GTGGTGGGGGGCGCGCTCAGTAGACGACAAGCGCGTGTCTGTCGGGCTCTTCGTAGACAA	1892
Db	1644	GTGGTGGGGGGCGCGCTCAGTAGACGACAAGCGCGTGTCTGTCGGGCTCTTCGTAGACAA	1703
QY	1893	GATCGTGTACACGAAGTCACTACGGGCGAGGGGCGAGGGAACGCCCATCGAAGCGCGC	1952
Db	1704	GATCGTGTACACGAAGTCACTACGGGCGAGGGGCGAGGGAACGCCCATCGAAGCGCGC	1763
QY	1953	TTTCGATCACGTTGGGCGGAAGCCGCCGACCGACGACGACGAAGAAAGACGACGCCACGAGCGCAC	2012
Db	1764	TTTCGATCACGTTGGGCGGAAGCCGCCGACCGACGACGACGAAGAAAGACGACGCCACGAGCGCAC	1823
QY	2013	GGAAGACGTAGCGCGC 2028	
Db	1824	GGAAGACGTAGCGCGC 1839	

```

RESULT 7
; US-09-788-297-28
; Sequence 28, Application US/09788297
; Patent No. US20020094516A1
; GENERAL INFORMATION:
; APPLICANT: Calos, Michele P.
; APPLICANT: Scilmenti, Christopher R.
; TITLE OF INVENTION: ALTERED RECOMBINASES FOR GENOME MODIFICATION
; FILE REFERENCE: 8400-0011
; CURRENT APPLICATION NUMBER: US/09/788, 297
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: altered
US-09-788-297-28

```

Query Match	87.0%;	Score 1767;	DB 10;	Length 1833;
Best Local Similarity	98.8%;	Pred. No. 0;		
Matches 1790;	Conservative	0;	Mismatches 20;	Indels 1;
				Gaps 1;

**OY**

**213** GATGTGCAGGTCACGGCGGCTTACGACCCTCAGTCGCCGCGACGCGAGAATAGCACTGC 272  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db	24	GGTGGACACGTTACGGCGGGTGTCTTACGACCGTCACTCGCGGAGCGCGAAATTTCAGCGC	83
QY	273	AGCAAGCCCCAGCGACACAGCGTAGCGCCAAAGACAAGAGCGCGGCGACCTTCAGCGCGA	332
Db	84	AGCAAGCCCCAGCGACACAGCGTAGCGCCAAAGACAAGAGCGCGGCGACCTTCAGCGCGA	143
QY	333	AGTCGAGCGCGAGCGGGGCGGGTTCAGGTTCTCGTGGGCAATTTCAGCGAAGCGCGGCGAC	392
Db	144	AGTCGAGCGCGAGCGGGGCGGGTTCAGGTTCTCGTGGGCAATTTCAGCGAAGCGCGGCGAC	203
QY	393	GTCGGCGTTCCGGACCGCGGAGCGCCCGGAGTTCGAACGCATCCTGAAAGCAATGCCGCGC	452
Db	204	GTCGGCGTTCCGGACCGCGGAGCGCCCGGAGTTCGAACGCATCCTGAAAGCAATGCCGCGC	263
QY	453	CGGGCGGCTCAACATGATCATTTGTCTATGACGTCGTCCGCTTCTCGCGCTGAAGGTCAAT	512
Db	264	CGGGCGGCTCAACATGATCATTTGTCTATGACGTCGTCCGCTTCTCGCGCTGAAGGTCAAT	323
QY	513	GGACGCGATTCCGATTTGTCTCTCGGAATTGTCTCGCCCTGGGCGTGACGATTGTTCCACTCA	572
Db	324	GGACGCGATTCCGATTTGTCTCTCGGAATTGTCTCGCCCTGGGCGTGACGATTGTTCCACTCA	383
QY	573	GGAAGGCGTCTTCGGCGCAGGGAACGTCATGGAACCTGATTACCTGATTTATCGCGCTCGA	632
Db	384	GGAAGGCGTCTTCGGCGCAGGGAACGTCATGGAACCTGATTACCTGATTTATCGCGCTCGA	443
QY	633	CGCGTCGCACAAGAAATCTTCGCTGAAGTCGGCGGAAGATTCTCGACACCGAAGAACTTCA	692
Db	444	CGCGTCGCACAAGAAATCTTCGCTGAAGTCGGCGGAAGATTCTCGACACCGAAGAACTTCA	503
QY	693	GCGGAATTGGGGGGGTACGTGCGGGGGAAGGGCGCTTACGGCTTCGAGCTTGTTCGGA	752
Db	504	GCGGAATTGGGGGGGTACGTGCGGGGGAAGGGCGCTTACGGCTTCGAGCTTGTTCGGA	563
QY	753	GACGAAGAGATCAACGCGCAACGGCCGAATGTCATGTGTCATCAACAAGTTAGCGCA	812
Db	564	GACGAAGAGATCAACGCGCAACGGCCGAATGTCATGTGTCATCAACAAGTTAGCGCA	623
QY	813	CTCGACCACTCCCTTACCAGGACCTTCGAGTTCGAGCCCGACGTAATCCGGTGTGTG	872
Db	624	CTCGACCACTCCCTTACCAGGACCTTCGAGTTCGAGCCCGACGTAATCCGGTGTGTG	683
QY	873	GCGTAGATCAAGACGCGCAAAACCTTCCCTTCAAGCCGGGCAAGTCAAGCCGCATTTCA	932
Db	684	GCGTAGATCAAGACGCGCAAAACCTTCCCTTCAAGCCGGGCAAGTCAAGCCGCATTTCA	743
QY	933	CCCCGGACGATCACGGGGCTTTGTAAGCGCATGACGCTGACGCCGTGCCGACCCGGGG	992
Db	744	CCCCGGACGATCACGGGGCTTTGTAAGCGCATGACGCTGACGCCGTGCCGACCCGGGG	803
QY	993	CGAGACGATTTGGGAAGAAGACCCTTCAAGCGCCTGGGACCCGGCAACCGTTATGCGAAT	1052
Db	804	CGAGACGATTTGGGAAGAAGACCCTTCAAGCGCCTGGGACCCGGCAACCGTTATGCGAAT	863
QY	1053	CCCTTCGGGACCCGCGTATTTGGGGCTTCCGCCGTGAGGTGATCTACAAGAAGACCGGA	1112
Db	864	CCCTTCGGGACCCGCGTATTTGGGGCTTCCGCCGTGAGGTGATCTACAAGAAGACCGGA	923
QY	1113	CGGACACCGCGACCAAGAAATGAGGGTTACCGCATTTCAAGCGCGGACCCGATCAGCGTCCG	1172
Db	924	CGGACACCGCGACCAAGAAATGAGGGTTACCGCATTTCAAGCGCGGACCCGATCAGCGTCCG	983
QY	1173	GCCGGTCGAGCTTGATTTGGGACCGGATCATCGAGGCCGCTGAGTGTATGAGCTTCAGGC	1232
Db	984	GCCGGTCGAGCTTGATTTGGGACCGGATCATCGAGGCCGCTGAGTGTATGAGCTTCAGGC	1043
QY	1233	GTGGTTGACGCGAGGGGGCGGCAAGGGGCTTCCCGGGGGCAAGCCATTCTGTCCGC	1292
Db	1044	GTGGTTGACGCGAGGGGGCGGCAAGGGGCTTCCCGGGGGCAAGCCATTCTGTCCGC	1103
QY	1293	CATGACAAGCTGTACTCGAGTGTGGCGCCGTCAATGACTTCGAAGCGCGGGGAAGAATC	1352
Db	1104	CATGACAAGCTGTACTCGAGTGTGGCGCCATCATGACTTCGAAGCGCGGGGAAGAATC	1163







```

1 FEATURE:
2 NAME/KEY: misc_feature
3 LOCATION: (4389)
4 OTHER INFORMATION: a, c, t, g, other or unknown
5 FEATURE:
6 NAME/KEY: misc_feature
7 LOCATION: (4400)
8 OTHER INFORMATION: a, c, t, g, other or unknown
9 FEATURE:
10 NAME/KEY: misc_feature
11 LOCATION: (4403)
12 OTHER INFORMATION: a, c, t, g, other or unknown
13 FEATURE:
14 NAME/KEY: misc_feature
15 LOCATION: (4411)
16 OTHER INFORMATION: a, c, t, g, other or unknown
17 FEATURE:
18 NAME/KEY: misc_feature
19 LOCATION: (4417)
20 OTHER INFORMATION: a, c, t, g, other or unknown
21 FEATURE:
22 NAME/KEY: misc_feature
23 LOCATION: (4424)
24 OTHER INFORMATION: a, c, t, g, other or unknown
25 FEATURE:
26 NAME/KEY: misc_feature
27 LOCATION: (4444)
28 OTHER INFORMATION: a, c, t, g, other or unknown
29 FEATURE:
30 NAME/KEY: misc_feature
31 LOCATION: (4455)
32 OTHER INFORMATION: a, c, t, g, other or unknown
33 FEATURE:
34 NAME/KEY: misc_feature
35 LOCATION: (4460)
36 OTHER INFORMATION: a, c, t, g, other or unknown
37 FEATURE:
38 NAME/KEY: misc_feature
39 LOCATION: (4465)
40 OTHER INFORMATION: a, c, t, g, other or unknown
41 FEATURE:
42 NAME/KEY: misc_feature
43 LOCATION: (4469)
44 OTHER INFORMATION: a, c, t, g, other or unknown
45 FEATURE:
46 NAME/KEY: misc_feature
47 LOCATION: (4478)
48 OTHER INFORMATION: a, c, t, g, other or unknown
49 FEATURE:
50 NAME/KEY: misc_feature
51 LOCATION: (4490)
52 OTHER INFORMATION: a, c, t, g, other or unknown
53 FEATURE:
54 NAME/KEY: misc_feature
55 LOCATION: (4500)
56 OTHER INFORMATION: a, c, t, g, other or unknown
57 FEATURE:
58 NAME/KEY: misc_feature
59 LOCATION: (4509)
60 OTHER INFORMATION: a, c, t, g, other or unknown
61 FEATURE:
62 NAME/KEY: misc_feature
63 LOCATION: (4578)
64 OTHER INFORMATION: a, c, t, g, other or unknown
65 FEATURE:

```

Query Match 2.5%; Score 50; DB 9; Length 15577;  
Best Local Similarity 55.9%; Pred. No. 0.28;  
Matches 95; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

[illegible]

```

RESULT 14
US-10-158-160A-14/c
; Sequence 14, Application US/10158160A
; Publication No. US200300059805A1
; GENERAL INFORMATION:
; APPLICANT: RAO, ERCOLE
; APPLICANT: RAPPOLD-HOERBRAND, GUDDRUN
; TITLE OF INVENTION: HUMAN GROWTH GENE AND SHORT STATURE GENE REGION
; FILE REFERENCE: 108351-00004
; CURRENT APPLICATION NUMBER: US/10/158,160A
; CURRENT FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: 09/147,699
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: PCT/EP97/05355
; PRIOR FILING DATE: 1997-09-29
; PRIOR APPLICATION NUMBER: 60/027,633
; PRIOR FILING DATE: 1996-10-01
; PRIOR APPLICATION NUMBER: EP/97100583.0
; PRIOR FILING DATE: 1997-01-16
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 32367
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-158-160A-14

```

Query Match	2.58;	Score 50;	DB 9;	Length 32367;
Best Local Similarity	55.9%;	Pred. No. 0.43;		
Matches	95;	Conservative	0;	Mismatches 75; Indels 0; Gaps 0;
QY 48	TACCTTGGATATATATATAATAATATCATTAATTAGTAGTAATATAATAATATCAATAT	107		
Db 7045	TTCTTATATGTATATATATGTTTATATATATTTTATATAGTATATATATATGTTTATATAT	6986		
QY 108	TTTTTCAAAATAAAGAATGTAGTATATAGCAATTCGTTTTCTGTAGTTTATAAGTGTG	167		
Db 6985	ATTATATATATACAAATATGTTGTATATATATTATATATGTTTATATATATATGTTTA	6926		
QY 168	TATATTTTAATTTATAACTTTTCTAATATATGACCAAAATTTGTTGATGT	217		
Db 6925	TATATTTATATATGTTTATATGTTTATATATATATATATATATATATATATAT	6876		

RESULT 15  
 US-10-239-676-117  
 ; Sequence 117, Application US/10239676  
 ; Publication No. US20030082609A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OLEK, Alexander  
 ; APPLICANT: PIEPENBROCK, Christian  
 ; APPLICANT: BERLIN, Kurt  
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
 ; FILE REFERENCE: 5013.1003  
 ; CURRENT APPLICATION NUMBER: US/10/239, 676  
 ; CURRENT FILING DATE: 2002-09-24  
 ; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
 ; DE 10019058.8  
 ; DE 10019173.8  
 ; DE 10032529.7  
 ; DE 10043826.1  
 ; PRIOR FILING DATE: 2001-04-06  
 ; 2000-04-06





**This Page Blank (uspto)**